University of Nottingham
UK | CHINA | MALAYSIA

Assessing the efficacy of Nicotine Replacement Therapy for smoking cessation during pregnancy

Ravinder Singh Claire, BSc (Hons) MPH

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

October 2020

Abstract

Background

Smoking during pregnancy is the leading modifiable risk factor for poor maternal and infant health outcomes. Pregnancy-related health problems associated with smoking during pregnancy include complications during labour, increased risk of miscarriage, premature birth, stillbirth and low birthweight. Despite this, around 12% of pregnant women in the United Kingdom (UK), 13% in the United States and 20% in France continue to smoke during pregnancy. A Cochrane review of 136 studies found that nicotine replacement Therapy (NRT) is proven to be effective amongst non-pregnant smokers, however a Cochrane review of eight studies found its efficacy in pregnancy to be uncertain. It is unclear whether we can ascertain a conclusion from this review as it may be subject to error due to repetitive testing, furthermore there may be insufficient power in the meta-analyses. Trial Sequential Analysis (TSA) is a method which could overcome these issues. This thesis provides an overview of TSA and applies the method to a systematic review of NRT use in pregnancy. This thesis also presents an alternative use for TSA, where it can be used for trial sample size estimation.

In most studies investigating NRT use for smoking cessation in pregnancy, women are instructed to discontinue use of nicotine patches if they have even brief smoking lapses. This is due to concerns that concomitant smoking and NRT use could increase exposure to nicotine and potentially more tobacco smoke toxins if they smoke heavily when using NRT. In 2014, the 'Study of Nicotine Patch in Pregnancy' (SNIPP) trial, a large randomised controlled trial (RCT) investigating NRT used in pregnancy for smoking cessation reported that it did not increase either smoking cessation rates or birth weights. This study was unique as participants were told that they

could continue using nicotine patches during smoking lapses. Using data from this trial, this thesis aims to explore whether concurrent smoking and NRT use resulted in changes in nicotine intake as well as smoking behaviour. This thesis also uses this trial to explore whether NRT use and changes in expired air carbon monoxide throughout pregnancy have an impact on birthweight.

Methods

Systematic review and meta-analysis

To determine the efficacy and safety of NRT for smoking cessation in later pregnancy, systematic review methods were used following standard Cochrane methods. The primary outcome was smoking cessation at the latest time point in pregnancy at which this was measured, and secondary outcomes were safety related. Meta-analyses were conducted where appropriate.

Trial Sequential Analysis

Trial sequential analysis was used to investigate whether there is sufficient evidence available to come to a firm conclusion on the efficacy of nicotine replacement therapy in pregnancy. Trial Sequential Analysis is a methodology that can be used in systematic reviews and meta-analyses to control random errors, and to assess whether further trials need to be conducted. We employ this method to the data from the systematic review, to assess whether there is sufficient evidence to conclude a clinically important treatment effect, no evidence of an effect, or lack of evidence.

This thesis goes on to explain an alternative use for Trial Sequential Analysis, where it can be used to estimate trial sample sizes for one or more trials investigating a behavioural smoking cessation intervention. We show

how data from a new, planned trial can be combined with data from the earlier trials using Trial Sequential Analysis to assess the intervention's effects. Using feasibility and pilot trials of a behavioural smoking cessation intervention, data are combined to estimate the sample size that one or more future RCTs would need to recruit, to provide a more decisive answer regarding intervention benefit.

Analysis of the SNIPP trial

The final study in this thesis used data from 402 women recruited to the SNIPP trial. Paired t-tests, linear regression, interaction tests, and within-individual variability analysis techniques were employed to answer the following questions: (1) does concurrent smoking and NRT use result in changes in nicotine, and other indicators of smoking intensity?; (2) do these changes differ between NRT or placebo patch use?.

Results

Systematic review and meta-analysis

Compared to placebo and non-placebo controls, there was low-certainty evidence that NRT increased the likelihood of smoking abstinence in later pregnancy (RR 1.37, 95% CI 1.08 to 1.74; I^2 = 34%, 9 studies, 2336 women). There was unclear evidence of an effect in placebo-controlled RCTs (RR 1.21, 95% CI 0.95 to 1.55; I^2 = 0%, 6 studies, 2063 women), whereas non-placebo-controlled trials showed clearer evidence of a benefit (RR 8.55, 95% CI 2.05 to 35.71; I^2 = 0%, 3 studies, 273 women).

Trial Sequential Analysis

The meta-analysis was not adequately powered to provide a strong conclusion, and TSA estimates that further placebo-controlled trials with

approximately 10,741 participants in total are needed to arrive at a firm conclusion.

Analysis of the SNIPP trial

(1) In the nicotine patch group, there was no change in saliva cotinine concentrations between baseline and 2-weeks post quit date (ratio of geometric means = 0.94, 95% CI=0.83 to 1.07; p=0.37, Bayes factor=0.15). However, there was a reduction in reported number of daily cigarettes smoked (mean difference -6, 95% CI's -7 to -5, p<0.001) and in CO concentrations (mean difference -3.0ppm, 95% CI's -4.2 to -1.9, p<0.001). (2) These changes were not significantly different from changes in the placebo group except for cigarette consumption which reduced more in the placebo group (p=0.046).

Conclusions

- NRT used for smoking cessation in pregnancy may increase smoking cessation rates in late pregnancy. However, this evidence is of low certainty, as the effect was not evident when potentially biased, non-placebo-controlled RCTs were excluded from the analysis.
- According to TSA, there is uncertainty regarding the efficacy of NRT use for smoking cessation during pregnancy compared to control, and further placebo-controlled trials are needed to arrive at a firm conclusion.
- Although TSA suggests more research is required for a firm conclusion, the general trend appears that NRT as it has previously been trialled, may not be effective for smoking cessation in pregnant women. Further trials should focus on what can be done differently in future. For example, using higher dose NRT or encouraging better adherence to treatment may produce more positive outcomes.

- Our findings suggest that when pregnant women use nicotine patches as part of a quit attempt, but they also smoke, they smoke less than they did before the quit attempt started. This means that their exposure to the toxic products of burnt tobacco is reduced.
- Despite having similar cotinine exposure to that from cigarette smoking, pregnant women who use nicotine patches and smoke, smoke less and exhale less CO, so their exposure to other tobacco smoke toxins is likely to be lower too.

Acknowledgements

This thesis was completed under the supervision of Professor Tim Coleman (Division of Primary Care, University of Nottingham), Professor Jo Leonardi-Bee (Division of Epidemiology and Public Health, University of Nottingham) and Dr. Ivan Berlin (Department of Pharmacology, Sorbonne Université, Faculté de medicine-Hopital Pitie-Salpetriere). Ravinder Claire's PhD is funded by the National Institute for Health Research (NIHR) Applied Research Collaboration East Midlands (ARC EM). The views expressed in this thesis are those of the author and not necessarily those of the National Institute for Health Research or the Department of Health and Social Care.

I am extremely grateful to my supervisors for their invaluable advice and patience during my PhD. Their guidance over the past four years has been exceptional, and without them I would not be where I am today. I hope one day we will be able to have a Skype meeting without any glitches! My sincere thanks also goes to Dr. Christian Gluud, for his technical support and his ever enthusiastic attitude to our meetings.

To all the comers and goers from 1502, it has been a pleasure sharing an office with you all. And to everyone I met in Paris and Cité Universitaire who made me feel at home, thank you for an unforgettable six months. Thank you to my friends, who have always been there when I have needed a pick me up or a laugh.

Finally, a very special mention to my family. My mum and dad, and my sister, Kiran – thank you for your unconditional love and support. You may not still fully understand my research, but none of this would have been possible without you. I hope I have made you all proud.

Table of contents

List of figures

List of tables

List of abbreviations

3HC	trans-3'-hydroxycotinine
ARC EM	Applied Research Collaboration East Midlands
ASQ	Ages and Stages Questionnaire, 3rd edition
BMI	Body Mass Index
BP	Blood Pressure
CCG	Clinical Commissioning Group
COPD	Chronic Obstructive Pulmonary Disease
CI	Confidence Interval
CINAHL	Cumulative Index to Nursing and Allied Health Literature
CO	Carbon Monoxide
FTCD	Fagerström Test for Cigarette Dependence
FTCQ-12	French Tobacco Craving Questionnaire, 12 items
ICC	Intracluster Correlation Co-efficient
IQR	Interquartile Range
IS	Information Size
IUGR	Intra-Uterine Growth Restriction
MD	Mean Difference
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NMR	Nicotine Metabolite Ratio
NRT	Nicotine Replacement Therapy
OR	Odds Ratio
RCT	Randomised Control Trial
RR	Risk Ratio
RRR	Relative Risk Reduction
SD	Standard Deviation
SGA	Small for Gestational Age
SIDS	Sudden Infant Death Syndrome
SNAP	The Smoking, Nicotine and Pregnancy trial
SNIPP	Study of Nicotine Patch in Pregnancy
TSA	Trial Sequential Analysis
UK	United Kingdom
USA	United States of America
WHO	World Health Organisation

List of publications

PhD related publications

Claire R, Coleman T, Leonardi-Bee J, Berlin I. Saliva cotinine concentrations in pregnant women who smoke and use nicotine patches. Addiction. 2019 Sep;114(9):1651-8.

Claire R, Chamberlain C, Davey MA, Cooper SE, Berlin I, Leonardi-Bee J, Coleman T. Pharmacological interventions for promoting smoking cessation during pregnancy. Cochrane Database of Systematic Reviews 2020, Issue 3. Art. No.: CD010078. DOI: 10.1002/14651858.CD010078.pub3.

Taylor L, **Claire R**, Campbell K, Coleman-Haynes T, Leonardi-Bee J, Chamberlain C, Berlin I, Davey MA, Cooper S, Coleman T. Fetal safety of nicotine replacement therapy in pregnancy: systematic review and meta-analysis. Addiction. 2020 Jul 4.

Claire R, Gluud C, Berlin I, Coleman T, Leonardi-Bee J. Using Trial Sequential Analysis for estimating the sample sizes of further trials: example using smoking cessation intervention. (Submitted to BMC Medical Research Methodology, revision two currently under review. Pre-print available at https://www.researchsquare.com/article/rs-35669/v2)

Associated publications

Hickson C, Lewis S, Campbell KA, Cooper S, Berlin I, **Claire R**, Oncken C, Coleman-Haynes T, Coleman T. Comparison of nicotine exposure during pregnancy when smoking and abstinent with nicotine replacement therapy: systematic review and meta-analysis. Addiction. 2019 Mar;114(3):406-24.

Chapter 1: Introduction

1.1 Smoking during pregnancy

The World Health Organisation (WHO) reports that approximately eight million deaths annually are caused by tobacco (1), and without further intervention tobacco could kill up to one billion people worldwide by the end of the century (2). All forms of tobacco are harmful and cigarette smoking is the most common form of tobacco use worldwide (1).

Smoking during pregnancy is a significant public health issue globally and is one of the leading preventable causes of poor health outcomes for mothers and their babies. Smoking tobacco during pregnancy exposes pregnant women to carcinogens, high concentrations of carbon monoxide (CO), nicotine and a multitude of other chemicals and heavy metals. The significant harms associated with smoking on both the mother and developing foetus, mean that smoking cessation and prolonged abstinence in pregnancy is critical for improving birth outcomes. In the following sections, the prevalence and harms of smoking in pregnancy will be discussed.

1.1.1 Epidemiology of smoking during pregnancy

It is estimated that 29 of 174 countries worldwide have a prevalence of smoking during pregnancy greater than 10%, and 12 countries have a prevalence of greater than 20% (3). The three countries with the highest prevalence of smoking are Ireland (38%), Uruguay (30%), and Bulgaria (29%) (3). Since the 1980's, high-income countries such as the Netherlands, Canada and Scotland have seen a decline in the prevalence of smoking in pregnancy from between 20% and 35% in the 1980s to below 10% in 2010 (4-6).

In England, the rates of smoking in pregnancy in England have been on the decline over the last 10 years (**Figure 1**). In 2006/07, smoking at time of delivery rates were 15.1%, and in 2011, the Tobacco Control Plan set an ambition to reduce smoking rates throughout pregnancy to 11% or less by the end of 2015 (7). This ambition was fulfilled in 2015/16 when smoking at the time of delivery rates declined to 10.6% (8). Whilst this decline is positive, recent data has shown that this rate has stagnated at 10.5% for 2016/17, with concerning variations by area (8).

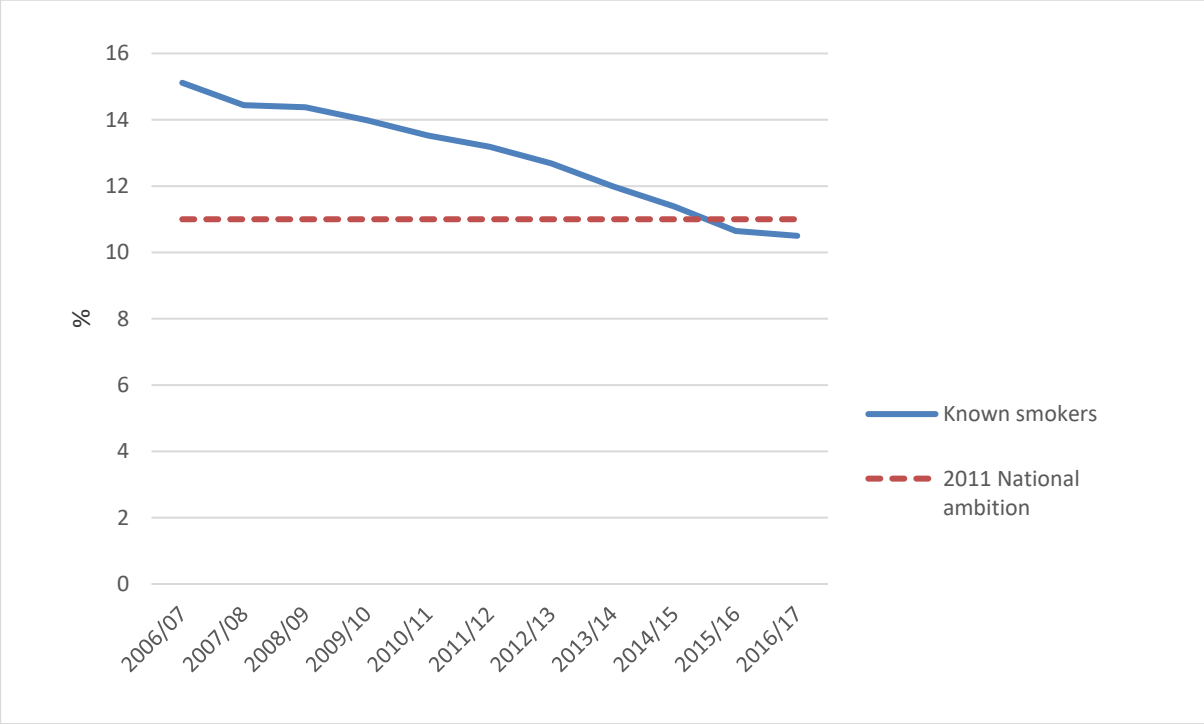

Figure 1 Women known to be smokers at time of delivery by year (8).

Smoking in pregnancy rates vary vastly throughout the UK, with NHS West London and NHS Richmond the Clinical Commissioning Groups (CCGs) having the lowest rates in the UK with rates of 2.3% and 2.5% for 2016/17 respectively (8). In contrast, the CCGs with the highest proportion of pregnant smokers were NHS Blackpool (28.1%) and NHS Hull (22.9%) - over 10% the national ambition (8). Less than half (104 of 209) of the CCGs

met the national ambition of reaching smoking during pregnancy rates of 11% or less, an increase of one from the previous year (8). Furthermore, all CCGs in the London commissioning region achieved the national ambition, whereas none of the 11 CCGs in the Cumbria and North-East commissioning region achieved this (**Figure 2**). As a result of the rates of smoking during pregnancy remaining stagnant over the last 2 years, and the disparity of rates throughout the UK, the latest tobacco control plan has made it an ambition to reduce the prevalence of smoking in pregnancy to 6% or less by the end of 2022 (9).

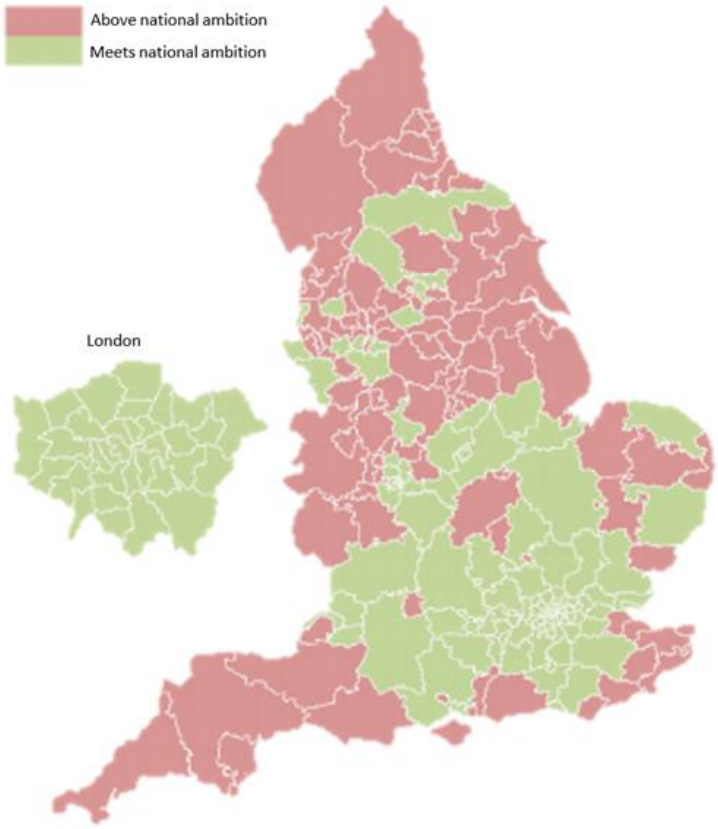

Figure 2 Women known to be smokers at time of delivery by CCG, compared with the national ambition (8).

1.1.2 Harms related to smoking during pregnancy

Smoking during pregnancy can cause a number of issues for both expectant mothers and their babies. Smoking during pregnancy is the leading

preventable cause of stillbirth, and babies that are born to mothers who smoke have a greater chance of being born underdeveloped or in poor health (9).

The rate of stillbirths in England and Wales in 2016, was the lowest it had been since 1992, at a rate of 4.4 stillbirths per 1000 total births (10). The UK's annual rate of reduction has been approximately 1.4% per year since 2000, however this is ranked 24th of 49 high income countries, and is small compared to the annual rate of reduction in the Netherlands and Poland (6.8% and 4.5% respectively) (11). Several studies have shown that smoking in pregnancy can increase the risk of stillbirth by approximately 30-50% (12-14).

As well as increasing the risk of stillbirth, smoking during pregnancy is also associated with increased perinatal and neonatal deaths (13), and a systematic review found that smoking increases the risk of miscarriage by approximately one quarter (15). A significant reduction in birthweight is also associated with smoking during pregnancy, this is most commonly defined as babies born <2,500g at ≥37 weeks gestation. Babies that are born to women who smoke through pregnancy weigh an average of 250g less than those from non-smoking mothers (16). Low birthweight is a risk factor for stillbirth, but is also associated to complications in later life, such as an increased risk of diabetes and cardiovascular disease (17). Smoking can also lead to babies being born small for gestational age (SGA). This is defined as a baby being born with a weight less than the standardised average for a given gestation. Smoking is considered to have a causal relationship with intra-uterine growth restriction (IUGR) and this affects the birth weight regardless of gestation (18).

Some studies have found that smoking during pregnancy can double the risk of preterm birth (19, 20), and is also now the principal risk factor for sudden infant death syndrome (SIDS) (21). SIDS is defined as the unexplained, sudden death of a child within the first year of life, and a meta-analysis found that prenatal smoking increased the risk of SIDS by 3 times (OR 2.94, 95% CI: 2.58-3.36) (16, 22). Associations have also been found with congenital abnormalities (such as orofacial clefts and musculoskeletal defects) and behavioural problems in later life (23, 24).

Smoking during pregnancy not only affects the baby, but also has a direct impact on the mother. Smoking is the leading preventable cause of morbidity and mortality, where approximately half of all smokers will die from a smoking related cause (25). Smoking prevalence amongst younger pregnant women and those in disadvantaged groups is considerably higher than older, more affluent women. Mothers in routine and manual occupations are 5 times more likely to have smoked during pregnancy, than women in managerial and professional occupations (26). Due to this disparity in prevalence rates, disadvantaged socioeconomic groups have higher rates of stillbirth, premature birth and low birthweight (16). Children that are born and grow up with a smoking parent, are more likely to become a smoker themselves, which continues the cycle of inequality (27). Smoking in pregnancy is a leading cause of health inequality and it is estimated to account for 38% of the inequality in stillbirth and 31% of the inequality in infant deaths (28).

1.1.3 Benefits of smoking cessation in pregnancy

Women are more likely to make an attempt at stopping smoking in pregnancy than at any other time in their lives. Despite this, 10.5% of women continue to smoke during pregnancy (29). It is estimated that

between 47% and 63% of women that do manage to quit during pregnancy, relapse to smoking within 6 months of delivery (30).

There are a number of benefits to both mother and child, if the mother quits smoking before pregnancy. A large population-based cohort study in Finland found that women who quit smoking in the first trimester of pregnancy had equal rates of stillbirth and preterm birth as non-smokers, and the prevalence of low birthweight and SGA outcomes were close to those of non-smokers (31). Smokers are approximately 4 times more likely to quit smoking if they use a stop smoking service (32). However, the number of pregnant women accessing specialist stop smoking services can be poor, with rates of engagement to these services as low as 12% of pregnant women who smoke (33).

1.2 Smoking cessation during pregnancy

Quitting smoking can reduce harm to pregnant mothers, their babies and members of their household. The following section will describe influences on smoking cessation in pregnancy and effective interventions that can be used to help pregnant women to stop smoking.

1.2.1 Influences on smoking cessation

An important factor for women to stop smoking is the realisation of their pregnancy. Smoking cessation rates are 3 times greater during the year of pregnancy (34), though few women quit smoking after the first trimester (35). This increase in cessation rate during pregnancy is likely to be because pregnant women are more likely to recognise the risks they pose to both the foetus and themselves, which can provoke a strong emotional response, motivating them to quit (36).

Smoking duration and age are both factors that have been associated with an increased number of quit attempts in pregnancy (37), whereas multiparity, increased nicotine dependence and having a partner that smokes are all factors that have been inversely associated with cessation (38). In addition, surveys have found that pregnant women with lower education and socioeconomic levels have decreased chances of cessation, whereas pregnant women who had a partner that did not smoke, started smoking when they were older, smoked fewer cigarettes or were primiparous were more likely to stop smoking (39).

1.2.2 Psychosocial interventions

Psychosocial interventions are defined as non-pharmacological strategies that use cognitive-behavioural, motivational and supportive therapies to help women to quit smoking, including counselling, health education, feedback, financial incentives, social support, and exercise (40). A Cochrane review, that included 120 RCTs and quasi-randomised studies, of psychosocial interventions for supporting women to stop smoking in pregnancy found that counselling interventions had a clear effect on cessation compared with usual care (RR: 1.44, 95% Confidence Interval [CI]: 1.19-1.73) and financial interventions also appeared to have a clear effect compared with an alternative, non-contingent incentive, intervention (RR: 2.36, 95% CI: 1.36-4.09) (40). Interventions that provided feedback with information about the foetal health status or measurements of by-products of tobacco smoking, also had a clear effect when compared with usual care and when combined with counselling (RR: 4.39, 95% CI: 1.89-10.21) (40). Health education (RR: 1.59, 95% CI: 0.99-2.55) and social support (RR: 1.21, 95% CI: 0.93-1.58) however, do not have a significant effect in stopping women from smoking during pregnancy (40). An

important factor for the use of psychosocial interventions during pregnancy is that there are no adverse outcomes associated with them (40).

1.2.3 Pharmacological interventions

Whilst some psychosocial interventions are successful in aiding pregnant women from stopping smoking, these interventions do not address nicotine addiction directly (41). Heavier smokers may require pharmacological treatments that substitute the nicotine delivery from smoking, to address addiction and metabolism of nicotine.

Pharmacological interventions that can help smokers quit include, nicotine replacement therapy (NRT), bupropion, varenicline or e-cigarettes (42). Whilst varenicline and bupropion have been successfully used for smoking cessation in the general population (43), there are currently no trials investigating varenicline in pregnancy, and the one trial investigating bupropion had recruitment issues and was only able to randomise 11 women (44). The lack of trials investigating varenicline and bupropion for smoking cessation during pregnancy is because there are currently no clinical guidelines that recommend their use, due to limited evidence for their safety during pregnancy (45). Additionally, use of bupropion and varenicline could expose the foetus to additional toxins found within these drugs, which is one reason why the study investigating bupropion struggled with recruitment (44).

A systematic review of electronic cigarettes for smoking cessation in the general population found evidence that e-cigarettes may work better than NRT (46). As yet, there are no published results investigating e-cigarette use to aid smoking cessation during pregnancy. However, both the WHO and Centers for Disease Control and Prevention advise that there is insufficient evidence to recommend e-cigarettes for smoking cessation in

adults, including pregnant women (47, 48). NRT is the most extensively studied pharmacological intervention for smoking cessation during pregnancy.

1.3 Nicotine Replacement Therapy

NRT is available in a variety of different forms including, transdermal patches, gum, spray and lozenges. Nicotine delivered by the gum, spray and lozenges offer brief, short-term doses of nicotine, whereas the patch acts over a longer-term (49). NRT works by substituting the nicotine inhaled in cigarette smoke with a medicinal form of nicotine. By using NRT, the toxins inhaled in cigarette smoke are avoided, whilst also relieving the withdrawal symptoms experienced when stopping smoking (49).

A Cochrane review of pharmacological interventions for smoking cessation included 9 studies in the review (50). This review identified the bupropion study discussed earlier, and 8 trials investigating NRT use for smoking cessation during pregnancy. The analysis of NRT in this review included a total of 2,199 pregnant smokers from 5 placebo-controlled studies (51-55), and 3 non-placebo-controlled studies (56-58), and found a borderline significant result for NRT used in pregnancy increasing smoking cessation rates by approximately 40% (RR: 1.41, 95% CI: 1.03-1.93) (59). However, a sub-group analysis of only placebo-controlled trials found that NRT was borderline not significantly effective in stopping women smoking during pregnancy (RR 1.28, 95% CI: 0.99-1.66) (50). The results from the Cochrane review show a clear disparity between the efficacy of NRT in the general population and the efficacy in pregnancy.

There could be a number of reasons for why there is a disparity between the efficacy of NRT in the general population and pregnancy. However, limited research about the factors that might influence pregnant women to

stop smoking when using NRT for cessation attempts has been conducted. One study found that women who were better educated had higher odds of stopping smoking at both one month into pregnancy and at delivery (60). Conversely, women who had higher baseline cotinine levels were inversely associated with cessation at both one month and at delivery (60). Adherence to NRT during pregnancy could be a potential factor to account for when determining the efficacy of NRT in pregnancy.

1.3.1 Adherence to NRT

In the general population, greater adherence with NRT has been found to be associated with increased odds of achieving cessation (49). Adherence to NRT in non-pregnant smokers appears to be high, with one study finding that 94% of smokers in a trial used NRT throughout their treatment period (61). By contrast, a Cochrane review found that only 7%-48% of pregnant women who received NRT, reported that they had completed a full course (50). Non-adherence of NRT for the prescribed period during pregnancy may restrict the efficacy of NRT (62). Adherence may affect the assessment of the efficacy and safety of NRT. Therefore, it is important to understand the causes of non-adherence and account for these in subsequent analyses.

The reasons for low adherence to NRT amongst pregnant smokers could be partially due to women's perceptions about the use of NRT and concerns that there could be potential foetal harms from nicotine (63). Another reason for low adherence could be due to an increase in nicotine metabolism during pregnancy (64). Pregnant women that smoke may not receive a high enough dose of nicotine from NRT to alleviate their cravings, therefore they may be unlikely to continue with the prescribed course.

Evidence suggests that, in the general population, increased adherence with NRT is associated with longer term smoking cessation. There is no such

evidence from studies in conducted in pregnancy. It is important to understand the possible causes of non-adherence with NRT in pregnancy, as well as the characteristics of pregnant women who are predominantly adherent. Future analyses should also investigate whether adherence to NRT in pregnancy is associated with smoking cessation.

1.3.2 Metabolism of Nicotine

Low adherence to NRT during pregnancy could be due to an increase in nicotine metabolism. Nicotine is primarily metabolized by the hepatic cytochrome CYP2A6 enzyme, with approximately 70-80% of nicotine metabolised via this pathway (65). The primary metabolite of nicotine is cotinine, which is then primarily metabolised to trans-3'-hydroxycotinine (3HC) (66). Measuring the ratio of nicotine to cotinine, or cotinine to 3HC is a way of measuring CYP2A6 activity, and both measurements are an indicator of nicotine metabolic rate (NMR) (66). Nicotine has a short half-life, whereas cotinine has longer, more stable half-life, meaning the measurement of cotinine to 3HC ratio is preferred (66). The cotinine to 3HC ratio can be ascertained effectively using saliva, blood or urine samples (67).

Changes in nicotine metabolism during pregnancy is a potential reason for non-adherence or reduced adherence to NRT. A combination of increased metabolic enzymes such as the CYP2A6 enzyme and increased liver blood flow are potential factors responsible for alterations of nicotine metabolism in pregnancy (68). One study found that clearance of nicotine and cotinine was 60% and 140% higher respectively, during pregnancy (69). This increase in NMR in pregnancy may mean that the fixed amount of nicotine derived from adhering to NRT might not be enough to suppress craving and withdrawal symptoms (69). It is important to ascertain whether current doses of NRT prescribed are sufficient enough to alleviate withdrawal

symptoms in pregnant women, and future studies should investigate differences between cotinine levels before pregnancy when smoking and during when using NRT.

Chapter 2: Aim and objectives

2.1 Aims

This thesis investigates the efficacy, safety and impacts on smoking intensity of Nicotine Replacement Therapy used for smoking cessation in pregnancy. As mentioned in **Section 1.3**, the last systematic review assessing the safety and efficacy of NRT was conducted in 2015 (59). This review found that NRT could be effective for smoking cessation in pregnancy (RR: 1.41, 95% CI: 1.03-1.93) (59). Since this review, a number of new studies may have been performed, therefore an update to this review is justified. Furthermore, it is unknown whether this meta-analysis is sufficiently powered to arise at a firm conclusion regarding the efficacy and safety of NRT for smoking cessation in pregnancy. If the meta-analysis is underpowered, it is unknown how many more studies are required to be able to come to a strong conclusion. To overcome this, a relatively new statistical methodology called Trial Sequential Analysis is introduced in this thesis. This method is appraised and there is a demonstration of how this can be utilised for planning trials in the context of smoking cessation interventions in pregnancy; as well as for supplementing meta-analysis in summarising data of existing trials of NRT for smoking cessation in pregnancy.

The overall aim is to investigate ways in which NRT use in pregnancy might be changed such that it has greater potential to be effective. These aims were investigated through objectives detailed below.

2.2 Objectives

I. To use conventional systematic review and meta-analysis to determine the efficacy and safety of NRT used during pregnancy for

smoking cessation in later pregnancy and after childbirth (**Chapter 5**).

II. To describe the limitations of meta-analysis and demonstrate how trial sequential analysis methodology can be used to supplement the findings of meta-analysis (**Chapter 3**).

III. To determine whether there is sufficient information in the meta-analyses presented for I above regarding the efficacy and safety of NRT for smoking cessation in later pregnancy (**Chapter 6**).

IV. To demonstrate how trial sequential analysis can alternatively be utilised to calculate trial sample size, using results from feasibility and pilot studies (**Chapter 4**).

V. To use the SNIPP trial to investigate the differences in indicators of smoking intensity in pregnant women when smoking before using NRT, and when using NRT and smoking concurrently (**Chapter 7**).

Chapter 3: Trial Sequential Analysis

3.1 Introduction

Systematic reviews and meta-analysis of randomised controlled trials (RCTs) are considered top of the hierarchy of evidence for decision making related to therapeutic interventions. To keep the evidence for decision making up to date, then systematic reviews, hence meta-analyses, require updating on a regular basis. However, the addition of data from more recent trials to the existing meta-analysis leads to significance testing being repeated, this increases the risk of random error and false-positive results. Trial Sequential Analysis (TSA) is a relatively new statistical method that has been developed to address these issues.

3.2 Aim

This chapter aims to discuss a background of reviews and meta-analysis, addressing biases and potential pitfalls of conducting a meta-analysis. This chapter will also discuss TSA and how this method can be used to supplement the findings of the meta-analysis. Additionally, criticisms of TSA will be addressed and the different types of outcome of TSA will also be discussed, using examples.

3.3 Combining sources of evidence

Healthcare decisions for both public policy and individual patients ought to be informed by the latest and best available research evidence (70). However, this can be challenging since there is a plethora of information available. In 2006, it is estimated that approximately 1,350,000 articles were published in over 24,000 peer-reviewed journals, and this number has been increasing year on year since (71, 72). This information can be found in both print and electronic media, from different countries and in a diverse range of languages. Furthermore the large amounts of information

generated by individual studies may be biased, methodologically flawed and can achieve conflicting results (73). It is unlikely that healthcare providers and policy-makers have the time, skills and resources to search, appraise and interpret this evidence, and then incorporate this into healthcare decisions (74).

3.3.1 Narrative review

Narrative reviews are the more traditional type of review found in medical literature, where experts summarise the evidence in their field from a theoretical or contextual standpoint (75). Narrative reviews provide readers with up-to-date information about a specific topic or theme.

The goal of a narrative review is to present an argument based on existing information aimed at an expert audience (77). Authors of narrative reviews must represent the evidence underpinning their argument (including but not limited to primary research), and demonstrate how the evidence has been collated to inform the reviews conclusions (77). Whilst traditional narrative reviews can be useful, the validity of a review depends on its methodological quality (75). Authors of narrative reviews may use subjective methods to collect and interpret data, and there is potential for authors to be selective in citing reports that support their ideas (76).

Narrative reviews are still commonly found in medical literature, but due to narrative reviews' risk of bias, systematic reviews are preferred for decision making (78). Systematic reviews involve the application of scientific strategies, in ways that limit bias, to the assembly, critical appraisal, and synthesis of all relevant studies that address a specific research question (78).

3.3.2 Systematic review

Systematic reviews are overviews of literature, undertaken by identifying, critically appraising and synthesising results of primary research studies using a strict, methodological approach, to answer a specific research question, thus making the available evidence more accessible to policy makers (79). This is done by framing a research question and then collating all empirical evidence that matches pre-defined inclusion criteria, which are set to answer the specific research question. Systematic reviews are based on strict, pre-specified, reproducible methods that aim to minimise bias, providing a greater reliability of findings (74). When conducted well, they can provide reliable estimates about intervention effects with defensible conclusions (80).

Systematic reviews are considered the pinnacle of evidence in the traditional hierarchy of evidence (81). This is because the specific methods in systematic reviews (**Figure 3**), limit bias and improve reliability and accuracy of conclusions (79). Systematic reviews can also be used to establish whether findings are consistent and generalizable across populations, settings and treatment variations (79). Where suitable, combining the results of several individual studies in a systematic review using statistical methods gives a more reliable and precise estimate of an intervention's effect than results from a single RCT, this is called meta-analysis (82).

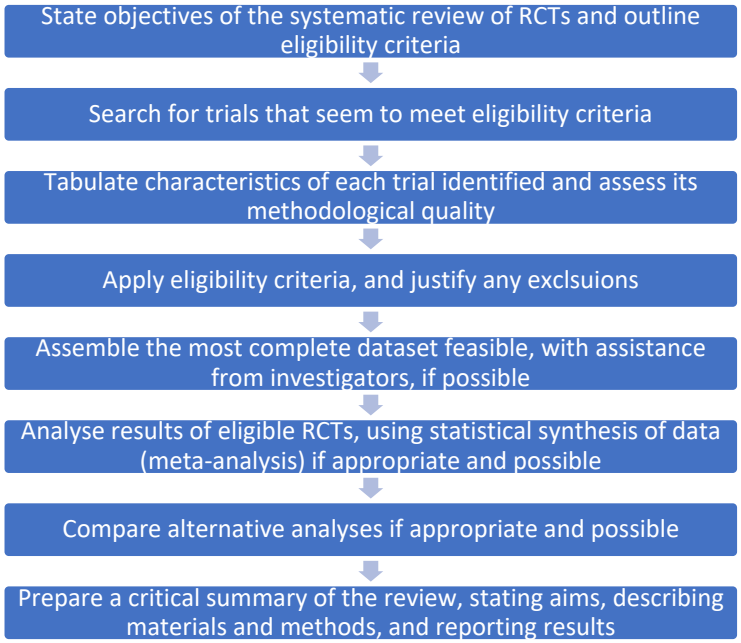

Figure 3 Methodology for a systematic review of randomised controlled trials.

3.4 Meta-analysis

Meta-analysis is the use of statistical methods to summarize the results of independent studies (82). A meta-analysis may be conducted following a systematic review, by pooling quantitative data from individual studies, and reanalysing them using recognised statistical methods (75). By combining the data from individual studies in a meta-analysis the overall sample size is increased, leading to a greater statistical power as well as more precision of the estimates of treatment effects (75).

Meta-analysis typically involves two stages, where the first stage calculates a measure of treatment effect with 95% confidence intervals for individual studies (83). The second stage of meta-analysis estimates an overall intervention effect as a weighted average of the individual summary statistics. When calculating this overall intervention effect, studies are weighted based on level of heterogeneity and the standard error of the

study, which takes into account the sample size and for dichotomous outcomes, the event rate in the two intervention groups (83).

Systematic review and meta-analysis are regarded as the most reliable sources of evidence as to whether an intervention should implemented into practice or further trials should be performed (84). Interventions are often recommended in clinical guidelines and implemented in clinical practice based on a meta-analysis showing statistical significance ($P<0.05$) (85). Additionally, meta-analyses published in the Cochrane Library are 57% more likely to be updated when they do not demonstrate statistical significance ($P \geq 0.05$) compared to those which do ($P < 0.05$), indicating that meta-analyses with statistically significant findings at the 5% level ($P < 0.05$) contribute to the decision to refrain from the updating of meta-analyses (86).

The findings included in a meta-analysis are typically presented in a forest plot, where an intervention effect size and 95% confidence interval is given for each study included (**Figure 4**). Each study is presented by a line and a solid square, where the lines represent the confidence intervals. The solid square represents the effect size for that individual study, and the area of the square is proportional to the study's weight in the meta-analysis. In meta-analysis, if all studies included were identical in terms of the methods and sample sizes used, one could simply calculate the mean of the effect sizes (87). However, it is rare to find all studies in a meta-analysis to be identical, therefore more weight is assigned to studies that carry more information and a weighted mean of the intervention effect is calculated. The pooled intervention effect and its 95% confidence interval are represented at the bottom of the forest plot by a diamond, where the lateral points indicate the confidence intervals for the estimate of the intervention effect.

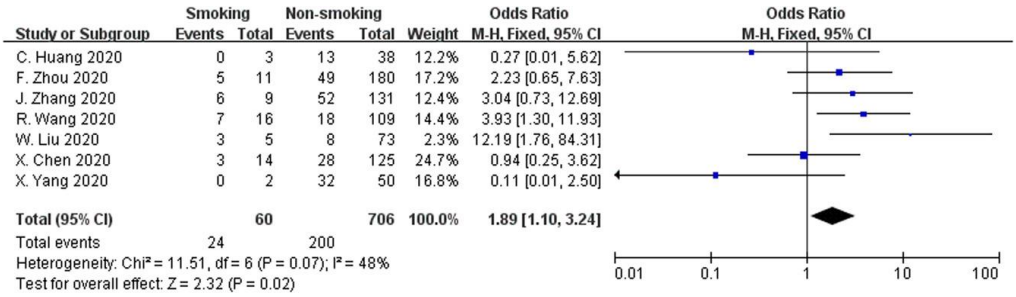

Figure 4 Example forest plot of smoking status and COVID-19 severity (88)

3.4.1 Fixed-effects and random-effects models in meta-analysis

The fixed-effect and random-effects models are two commonly used models used in meta-analysis. These models make different assumptions about the nature of studies included, and thus lead to differing methods for assigning weights (87).

The fixed-effect model assumes that the true intervention effect size is the same across all studies, and the pooled estimate of effect is an estimate of this common intervention effect size (87). Therefore, it is assumed that the sole reason the intervention effect size differs between studies is due to sampling error (chance). The weighting typically used in this model is based on the inverse variance of the individual studies, thereby assigning less weight to smaller studies. A limitation of this model is that a meta-analysis which only includes one large study and several relatively small sized studies would give the vast majority of the weight to the large study (87); thus the result for the meta-analysis would be very similar to the result of the large study. The fixed-effect model assumes that studies are identical in design and population and hence there is little variation between them; however,

this assumption may not be true for many systematic reviews of healthcare interventions. When studies are included in a systematic review, the inclusion criteria set means that studies are similar enough so that a single estimate of the intervention can be determined. However, this does not mean that all of these studies have to be identical, in the sense that the true intervention effect size is exactly the same for all studies (87).

Systematic reviews addressing a clinical question draw together several studies. Whilst these studies are only included if they match set inclusion criteria, it is inevitable that there will be some element of diversity between studies. Studies may differ in design, participants, interventions exposures or outcomes; this is called heterogeneity (89). The random-effects model assumes that the true intervention effect varies between each study, and the studies included in the meta-analysis represent a random sample of all of the potential intervention effects that could have been observed in individual studies, thus the pooled intervention effect is an estimate of the mean of the effects (87). In the random-effects model, heterogeneity is modelled within the weightings, so that the weights assigned to each study is a combination of both the standard error of the individual study and an estimate of heterogeneity between studies. The effect of this is that as the estimate for heterogeneity increases, the weights will be more evenly distributed between the studies – i.e. smaller studies are given more relative weight, and larger studies are given less relative weight (87).

3.4.2 Cumulative meta-analysis

In 1992, Lau *et al.* (90) developed a new technique for updating meta-analyses whenever a new study is published, thus enabling the evaluation of the pooled intervention effect as a continuum. This is known as cumulative meta-analysis. The advantage of this method over conventional

meta-analysis is that by updating a meta-analysis routinely, the benefit or harm of an intervention can be identified as early as possible (90). Alternatively it can be used to justify commencement of new trials, or to question whether further trials should be carried out (91).

3.4.3 Publication bias in meta-analysis

The intention of a meta-analysis is to summarise the intervention effects from all available studies; however, this may not be possible; for example, where some studies are not identified from literature searching. Electronic databases such as MEDLINE and PUBMED do not contain all medical journal papers, and solely searching these would be insufficient to capture all studies addressing a specific research question. Furthermore, studies are less likely to be published if the intervention effect was not statistically significant, this is known as publication bias (92).

3.4.4 Random error in meta-analysis

The result from a meta-analysis is usually deemed positive or negative based on a test statistic, communicated with a p-value or confidence interval (93). Meta-analyses can sometimes yield false-positive (type I error) or false negative (type II error) results (94). Type I errors occur when chance (random error) is the cause of a positive meta-analysis result, rather than due to a 'true' intervention effect. Conversely, some negative meta-analytic results may be due to lack of statistical power and precision, yielding a type II error (94).

Meta-analysis methods do not consider the amount of the available evidence in relation to the required sample size (86, 95, 96). The reliability of a statistically significant intervention effect generated by meta-analysis is often overvalued, particularly where sparse data (e.g. number of events and

participants) or repetitive analyses (type I errors) are seen (74, 93, 97, 98). In meta-analyses with many study participants and studies with similar findings, test statistics and intervention effect estimates will tend to converge towards the true intervention effect (93). **Figure 5(A)** and **(B)** demonstrate examples of convergence in test statistics. In both figures, interpretation of statistical significance are inaccurate in early studies, but eventually converge toward the 'true' side of statistical significance as subsequent studies are included (93).

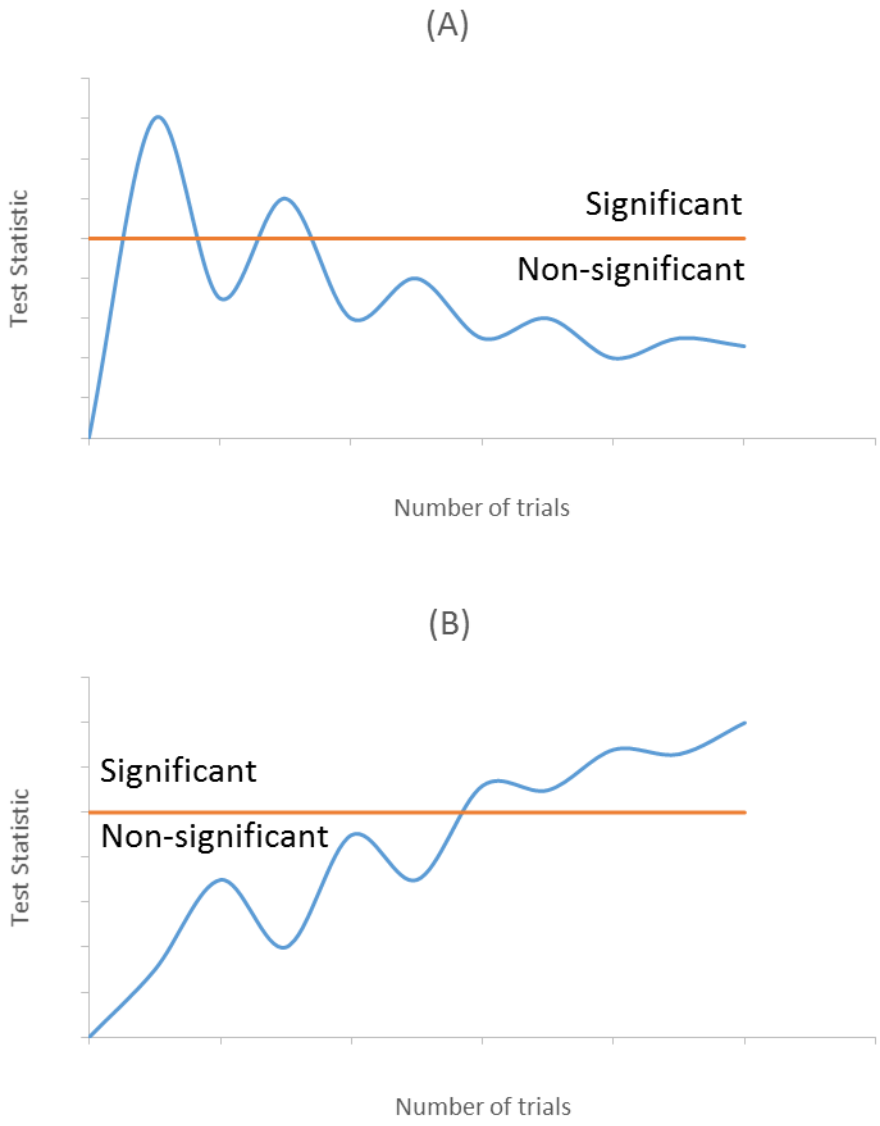

Figure 5 Examples of convergence in test statistics as studies are included and followed to an outcome measure in two meta-analyses A and B (93).

Random error and imprecision only cause problems if statistical tests are performed at stages where the extent of the random error is substantial enough to yield spurious statistical conclusions (93). For example, in **Figure 5(A)**, significance testing during the two peaks in early trials would lead to a false positive result. Similarly, in **Figure 5(B)**, early significance testing would have led to a false-negative conclusion.

The likelihood of observing a false-positive or false-negative result is greater with an increasing number of statistical tests performed on accumulating data. This is known as 'multiplicity due to repeated significance testing' (99). It is important for meta-analyses to minimise the risk of making a false-positive or false-negative conclusion. Pooled intervention effects in meta-analysis are usually assessed using P-values, and meta-analysts must decide on the threshold at which a P-value is sufficiently small to justify a 'positive' conclusion or the threshold below which a P-value is considered statistically significant (93). Deciding on a threshold involves a trade-off between the risk of observing a false-positive and false-negative result.

When significance tests are performed with few studies in a meta-analysis, or performed multiple times, there is an increase in the risk of observing a false result. Therefore, interpretations about statistical significance should be made in relation to the strength of evidence. That is, the total number of participants, observed number of events (for dichotomous outcomes), as well as the impact of multiplicity (100).

3.4.5 Limitations of meta-analysis

Meta-analyses aim to discover the benefit or harm of an intervention as early and as reliably as possible, as a result they tend to be updated when new studies are published (101). In previous years, reviewers which published their reviews in the Cochrane Library were required to update their systematic reviews at least once every two years, however they are now updated based on priority (74). When meta-analyses are updated, they are subjected to repeated significance testing, which has been shown to increase the risk of type I error (102) by between 10% and 30% (99). In practice, this means that between 1 and 3 out of 10 treatments implemented based on meta-analysis results are likely to be inappropriate.

Another limitation of conventional meta-analysis methods is that they do not consider the amount of the available evidence, and the reliability of a statistically significant intervention effect is often overvalued, irrespective of the number of events and participants (74, 93). In addition, intervention effects that don't show statistical significance are seen as unreliable, and it is assumed that more evidence is required (103).

A criticism of cumulative meta-analysis is that there are no guidelines for assessing whether statistical evidence is conclusive or not, other than the nominal P-value calculated from a meta-analysis after a new trial is added (91). This P-value does not fully take into account the amount of information or the number of participants included in the analysis (91). Additionally, there is an increased risk of random error in cumulative meta-analysis.

There is no way to differentiate between an underpowered meta-analysis and a true finding of an intervention being 'ineffective'. However, it is imperative that a conclusion as to whether an intervention is truly ineffective or truly effective is made as soon as possible after studies are completed, in order to guide investigators' decisions as to whether further studies could be informative or not (93). TSA is a method that can overcome this issue by distinguishing whether meta-analyses provide evidence for either beneficial or harmful intervention effects, lack of effect (futility), or insufficient evidence for evaluation of the intervention effect (93, 104).

3.5 Trial Sequential Analysis

As discussed in **Section 3.4**, meta-analyses aim to discover the benefit or harm of an intervention as early and as reliably as possible. As a result, they tend to be updated when new studies are published (101). When intervention evaluation has just begun and only few, smaller trials are available, meta-analyses may be conducted on sparse amounts of data and

their findings are therefore at high risk of random error (105). As meta-analyses are updated they are subjected to repeated significance testing, which increases the risk of type I errors (102). When there are few data available, TSA resolves these issues by having stringent thresholds for assessing statistical significance, using monitoring boundaries. Monitoring boundaries also take into account the volume of significance testing which has been undertaken through adjusting the thresholds that are used to define whether or not results are considered statistically significant (93).

TSA is also able to assess when an intervention has an effect smaller than what would be considered clinically minimally important (93). Futility boundaries, originally developed for interim analysis in RCTs, can be estimated and used to provide a threshold below which an intervention would be considered to have no clinically important effect (102). Thus, performing further trials is considered futile as the intervention does not possess the postulated clinically minimally important effect (93).

In TSA, when neither the monitoring boundaries nor the futility boundaries are crossed, further information is required. TSA can also inform how much more information is required to provide a conclusive answer regarding the effect of the intervention versus its comparator – this is the distance between the accrued information and the required information.

TSA can be used on all meta-analyses, and uses an approach that is analogous to the interim analysis of single RCTs developed by Lan and DeMets (106). In TSA of meta-analysis, trials are included in chronological order, and interim analysis is performed on them relative to the required number of participants for conclusive findings regarding intervention efficacy (information size). If the studies accrued in the TSA does not reach the information size, the uncertainty of the intervention effect will increase

(104). The more participants included, the smaller the uncertainty. When the required information size has not been reached, the threshold for significance is adjusted. The fewer participants in the TSA, the lower the significance level is in order to reliably assess the uncertainty of the estimated intervention effect (104). **Figure 6** shows a labelled example output from a TSA report.

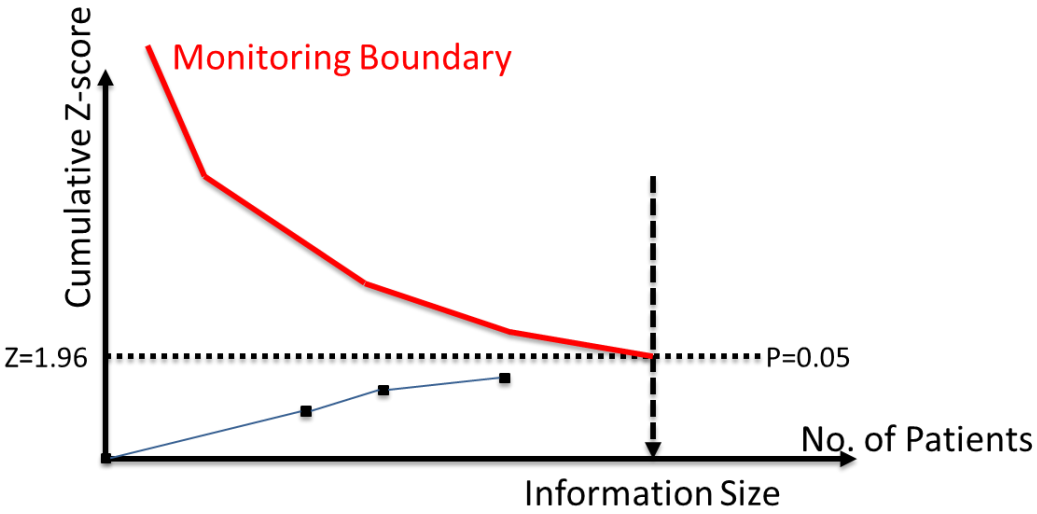

Figure 6 An example output from a TSA report. Each individual square represents a different study in chronological order. The blue line is the cumulative z-line, and represents the significance. The horizontal dotted line represents the conventional test boundary (p=0.05). The red line is the adjusted monitoring boundary – the cumulative z-line will cross this if there is evidence of an effect. The dashed vertical arrow represents the information size – this is the required number of patients needed to come to a firm conclusion.

3.5.1 Information size

If all available studies are included, systematic reviews and meta-analyses are considered the best available evidence, because power and precision of the estimated intervention effects are increased in meta-analyses compared to using a single study alone (74). However, this does not necessarily mean that the available evidence is either sufficient or strong enough to be able to provide a conclusion.

Evidence suggests that intervention effects and P-values based on few events and participants are unreliable (107). Approximately one quarter of conventional meta-analyses with a small number of participants and events may falsely pronounce the estimated intervention effects as statistically significant (108). Furthermore, positive large pooled intervention effects observed in early meta-analyses, tend to dissipate as more evidence is gathered (108-110).

For individual trials, an estimation of the required sample size is performed to ensure the number of participants included is enough to detect or reject a minimum clinically important effect size (104). For dichotomous outcomes, such as death, the sample size estimation is based on the expected proportion of deaths in the comparator group, the expected relative risk reduction of the intervention, and the selected maximum risks of both type I and type II errors (101). Similarly, for meta-analyses to produce adequately powered findings regarding an intervention effect, sufficient numbers of participants need to be included. This number is referred to as the 'required information size' (also known as 'optimal information size' and 'meta-analytic sample size') (102, 105, 111). The required information size can be estimated using similar parameters as those used in sample size estimation for a single study. If it is applicable to consider random-effects model for assessing the intervention effect size, then an adjustment for between-study heterogeneity, measured by diversity (D^2), is needed (104). Heterogeneity between studies is likely to be observed in meta-analyses due to the magnitude of the intervention effect varying when used in different study populations, in studies with different methodological characteristics, or due to variations in the intervention itself (96). Thus, sample size estimations need to be increased to allow for this between-trial heterogeneity (104).

In TSA, studies are chronologically ordered, and interim analyses are conducted as each study is added. In a TSA where the 'required information size' has not been reached, the threshold for statistical significance is inflated to account for sparse data and multiple testing of the interim analyses using monitoring boundaries; thus, the 95% confidence interval is not providing coverage of the real uncertainty and the cut-off for determining statistical significance is below the usual nominal figure of 0.05 (104).

3.5.2 Significance testing with inadequate information size

As discussed in **Section 3.4.5**, meta-analyses are subjected to repeated significance testing when they are updated, increasing the risk of type I error. A resolution to solve this problem is to adjust the thresholds which are used to define whether or not results are considered statistically significant (93). **Figure 7** demonstrates an example of a meta-analysis where false-positive results are avoided using monitoring boundaries adjusting the threshold for statistical significance.

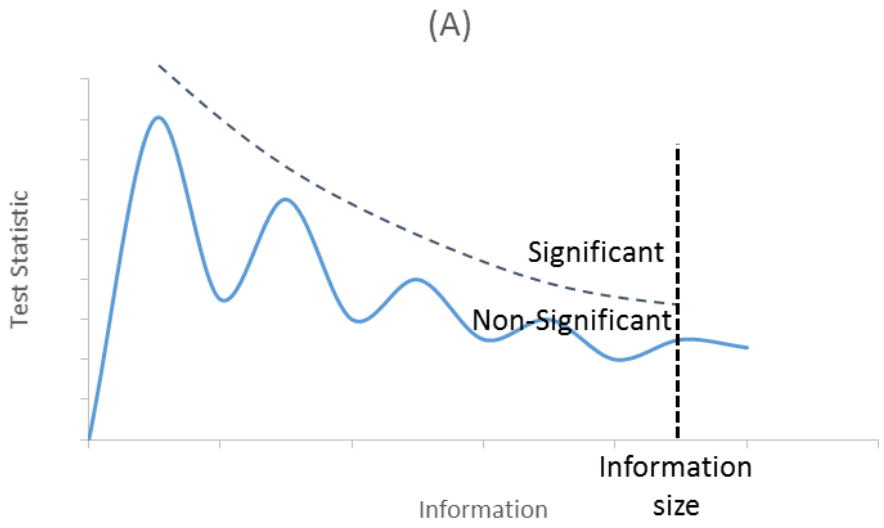

Figure 7 Examples of significance threshold adjustment (stipulated monitoring boundaries) (93).

3.5.3 Futility testing with inadequate information size

Meta-analyses often influence future research. Before developing future studies, investigators require an accurate summary of the most up to date information. If a meta-analysis has found that an intervention has no significant effect, it is important to ascertain how valid this finding is and to be able to exclude the possibility that the intervention is really effective but meta-analysis findings have arisen due to a lack of power (93). Using the TSA approach, unless an appropriate information size has been reached, when an intervention is found to have no effect, such a finding would be considered to be due to lack of power (93). Without using an approach such as TSA however, one would be unable to differentiate between an underpowered meta-analysis and a true 'ineffective' finding. However, it is imperative that a conclusion as to whether a treatment effect isn't as large as expected, is made as soon as possible in order to prevent investigators spending resources on unnecessary further studies (93). Alternatively, the anticipated intervention effect can be re-evaluated, and further research can

be designed to investigate whether there is evidence of a smaller, but still clinically significant, intervention effect size.

TSA is able to assess when an intervention has an effect smaller than what would be considered minimally important as early as possible (93). Futility boundaries, originally developed for interim analysis in RCTs, are created and used to provide a threshold which an intervention would be considered to have no effect (102). In a sufficiently powered meta-analysis, if an intervention is truly an improvement compared to the comparator, the test statistic would be expected to fluctuate around an upward sloping straight line, eventually yielding statistical significance (93). In a meta-analysis with fewer events and participants, obtaining a statistically significant result is unlikely due to lack of power. As further studies are introduced, the risk of getting a negative finding due to chance is reduced. Futility boundaries are a set of thresholds that reflect the uncertainty of obtaining a chance negative finding in relation to the number of participants (93).

If a test statistic is above the futility boundary, the test statistic may not have returned statistical significance due to lack of power, however there is a chance that that a statistically significant effect will be found before the meta-analysis exceeds the information size (93). If a test statistic is below the futility threshold, the test statistic is so low that the likelihood of a significant effect being found becomes negligible. At this point, performing further studies is futile as the intervention does not possess the postulated effect (93). **Figure 8(A)** illustrates an example of a meta-analysis where the intervention is not superior to the comparator. The test statistic crosses the futility boundary before the required information size is passed. **Figure 8(B)** demonstrates an example of a meta-analysis where the intervention is statistically significantly superior to the comparator. Here the test statistic

stays above the futility boundary and also yields statistical significance by crossing the monitoring boundary (93).

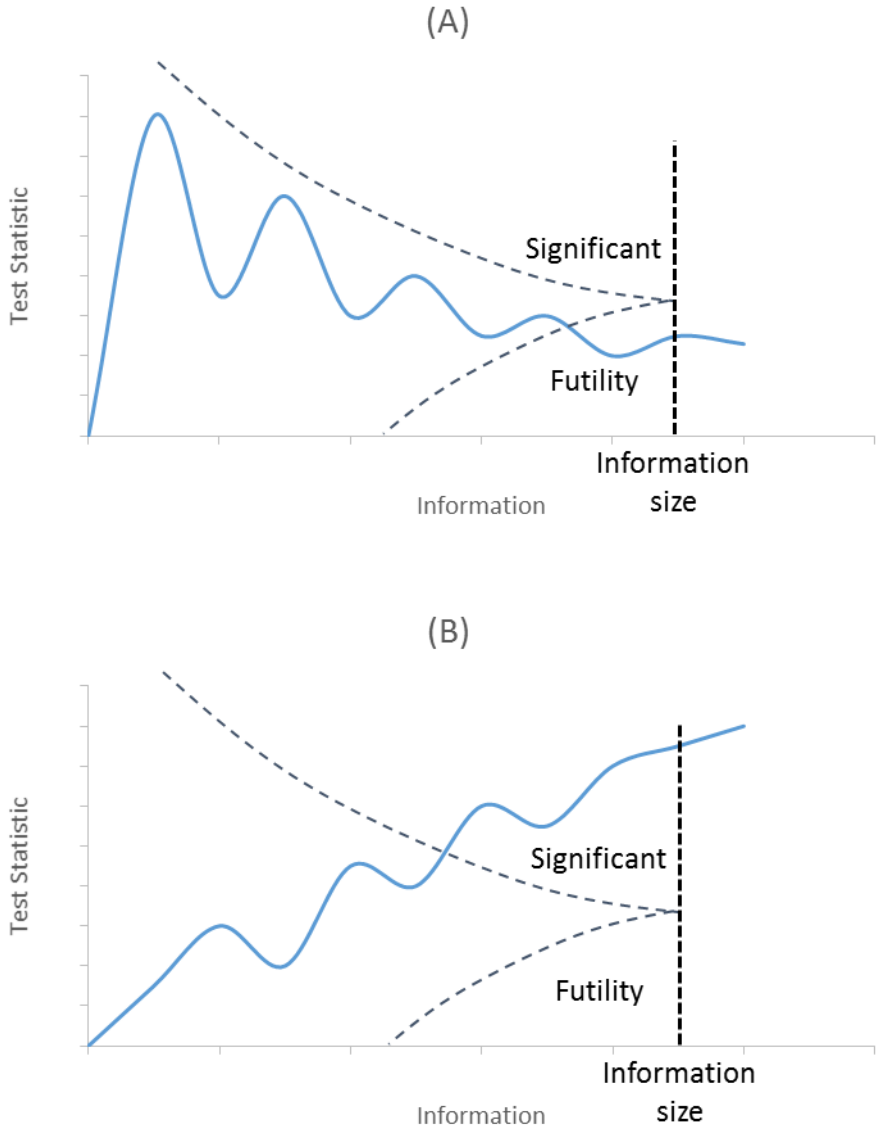

Figure 8 Examples of futility boundaries where the experimental intervention is not superior to the control intervention (and unnecessary trials may have been conducted) **(A)** and where the experimental intervention is statistically significantly superior to the control intervention (again where unnecessary trials may have been conducted) **(B)** (93).

3.5.4 Example results from trial sequential analysis

This section aims to illustrate the various results that can be yielded from using TSA. **Figure 9** shows a TSA of a meta-analysis comparing two treatments A and B where the Y-axis signifies the cumulative Z-score and

the x-axis signifies the cumulative number of participants included in the meta-analysis. In this TSA, the information size required is 855, however the cumulative Z-score crosses the monitoring boundary after two studies have been included. Even though the required information size has not been reached, it can be concluded that intervention B has a greater effect than intervention A and perhaps the intervention effect is larger than the anticipated effect (112). Thus, there is sufficient evidence to provide a firm conclusion and further studies based on this research question are not required. Screenshots of the TSA software to demonstrate the inputs required can be found in **Appendix D**.

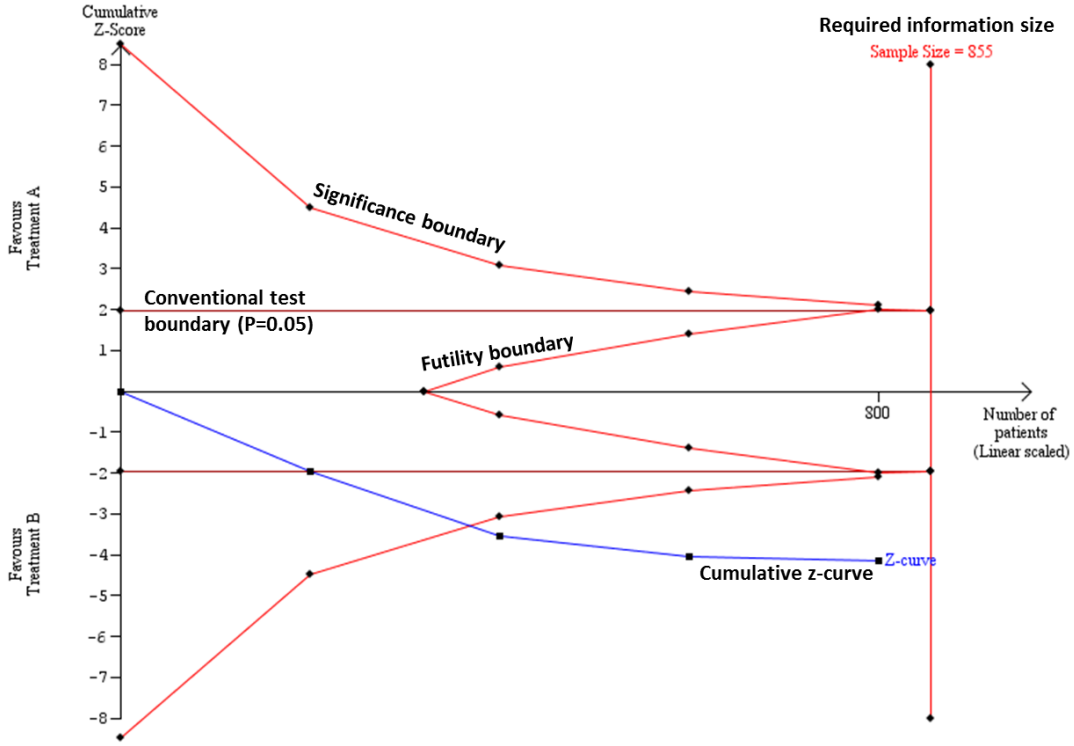

Figure 9 Example TSA showing the cumulative Z-score crossing the monitoring boundary, but information size has not been reached.

The cumulative Z-score crosses both the monitoring boundary as well as the required information size in **Figure 10**. Again, the conclusion is that intervention B is superior to intervention A and that the intervention effect is larger than the anticipated effect. Similar, to **Figure 9**, further studies

are not required. However, three studies had been conducted after a firm conclusion was already determined; therefore if a TSA had been conducted earlier, perhaps these studies could have been avoided and resources been better placed elsewhere.

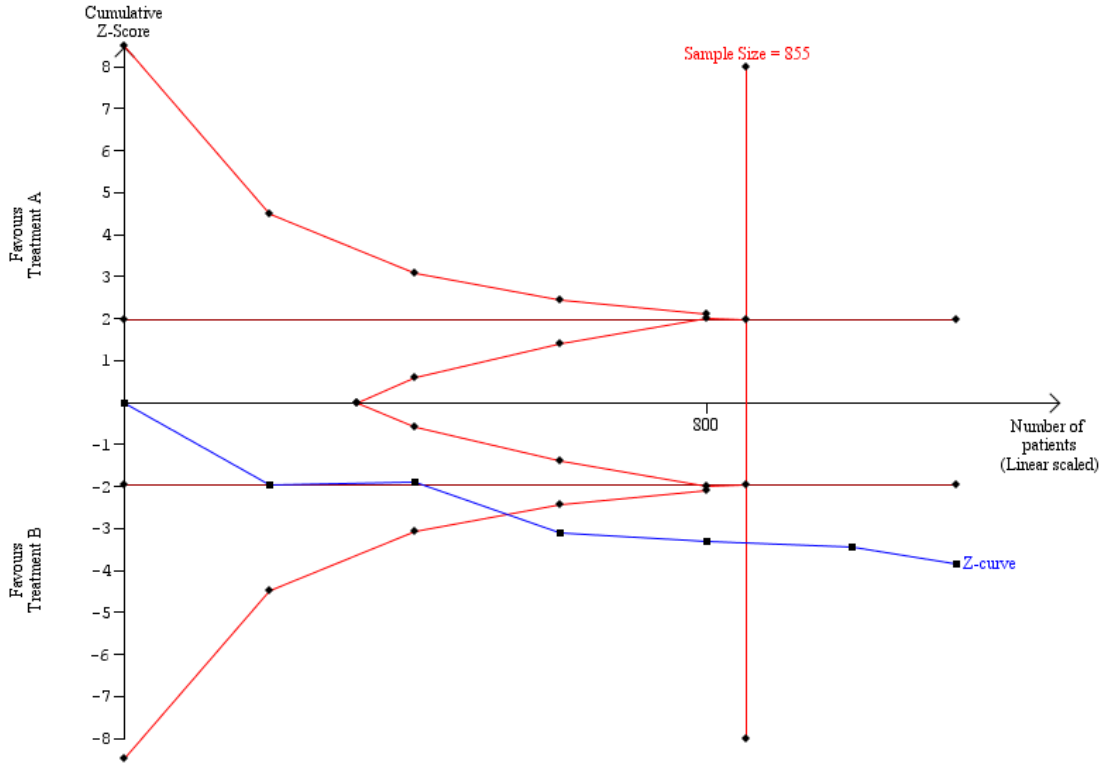

Figure 10 Example TSA showing the cumulative Z-score crossing the monitoring boundary, and information size has been reached.

In meta-analysis, it is important to understand whether a non-significant result is truly down to a lack intervention effect, or whether this result is due to lack of statistical power. TSA enables this differentiation. In **Figure 11** the Z-curve does not cross either the monitoring boundary or the conventional test boundary (P=0.05). However, a sample size of 1143 was not sufficient to reach the required information size (2144), therefore more studies are required.

Whereas in **Figure 12** the cumulative Z-curve crosses the futility boundary. When this occurs it can be inferred that the intervention effect is smaller

than what would be considered minimally clinically important to participants (93). **Figure 12** also demonstrates that the futility boundaries were crossed after the fifth study was included in the TSA, suggesting that the sixth study was not required.

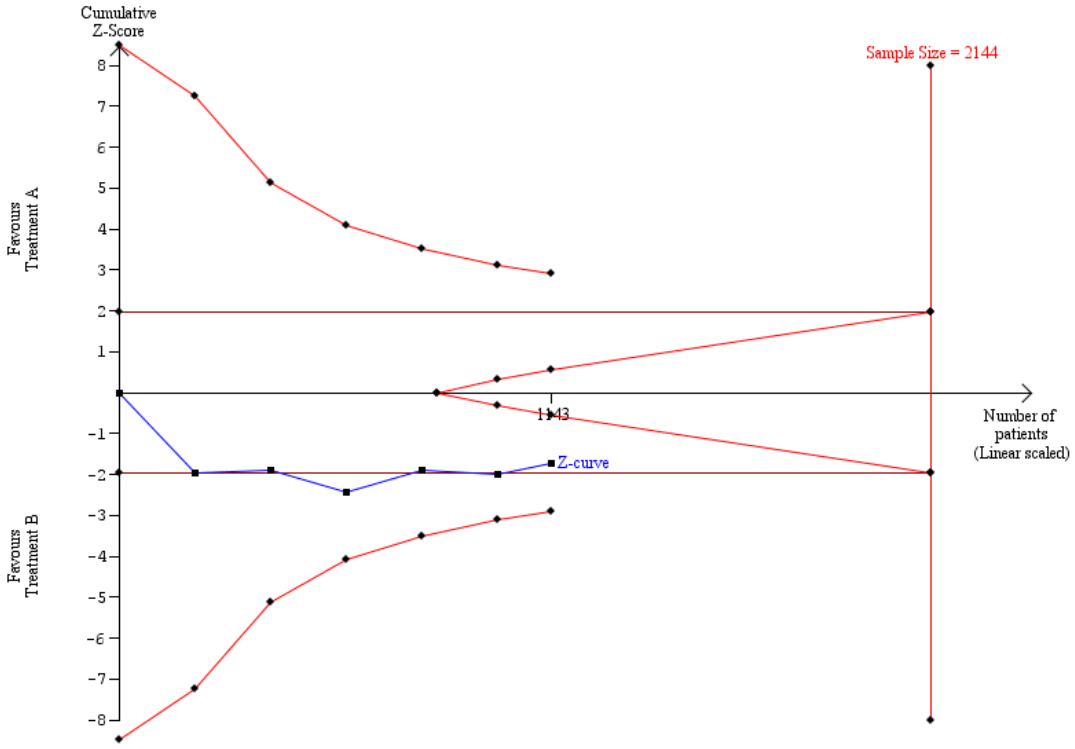

Figure 11 Example TSA showing the cumulative Z-score not crossing the monitoring boundary or the conventional test boundary, and information size has not been reached.

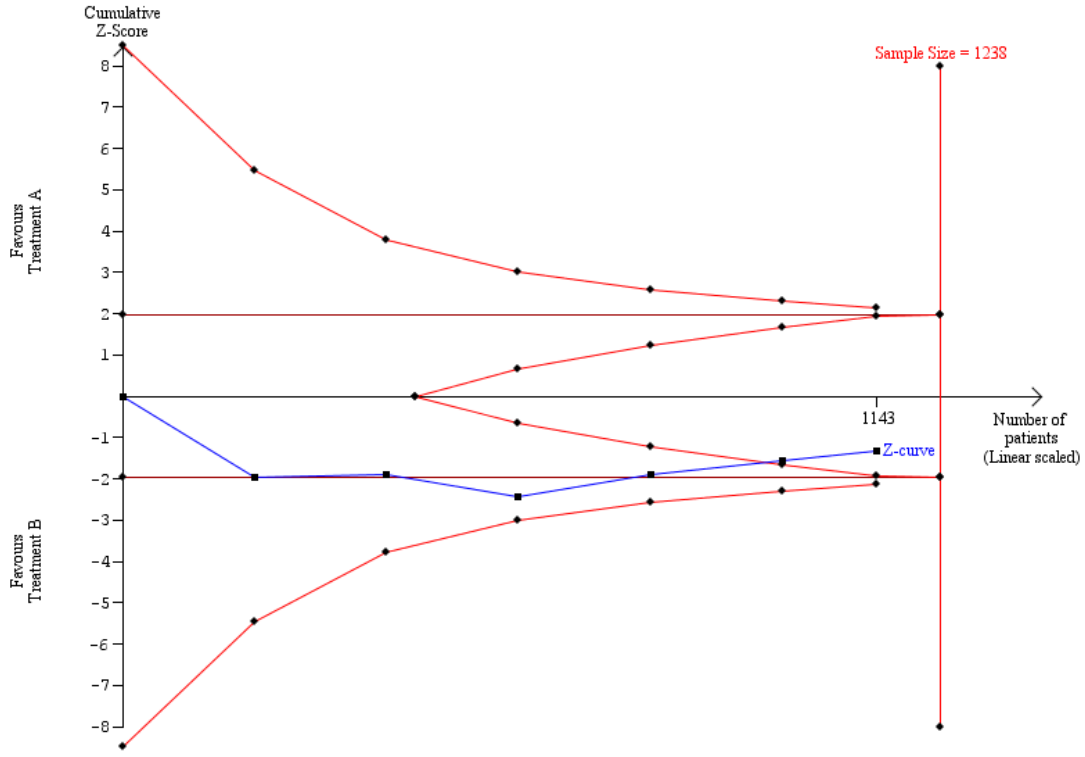

Figure 12 Example TSA showing the cumulative Z-score crossing the futility boundary.

Figure 13 shows the cumulative Z-score crossing the conventional test boundary, however the monitoring boundary has not been crossed. Futility boundaries have not been crossed, suggesting that there could be a significant intervention effect but the required information size has not been reached, deeming the meta-analysis inconclusive with more studies being required before a firm conclusion can be made. Specifically, further studies with a total of approximately 930 participants are required to come to a firm conclusion.

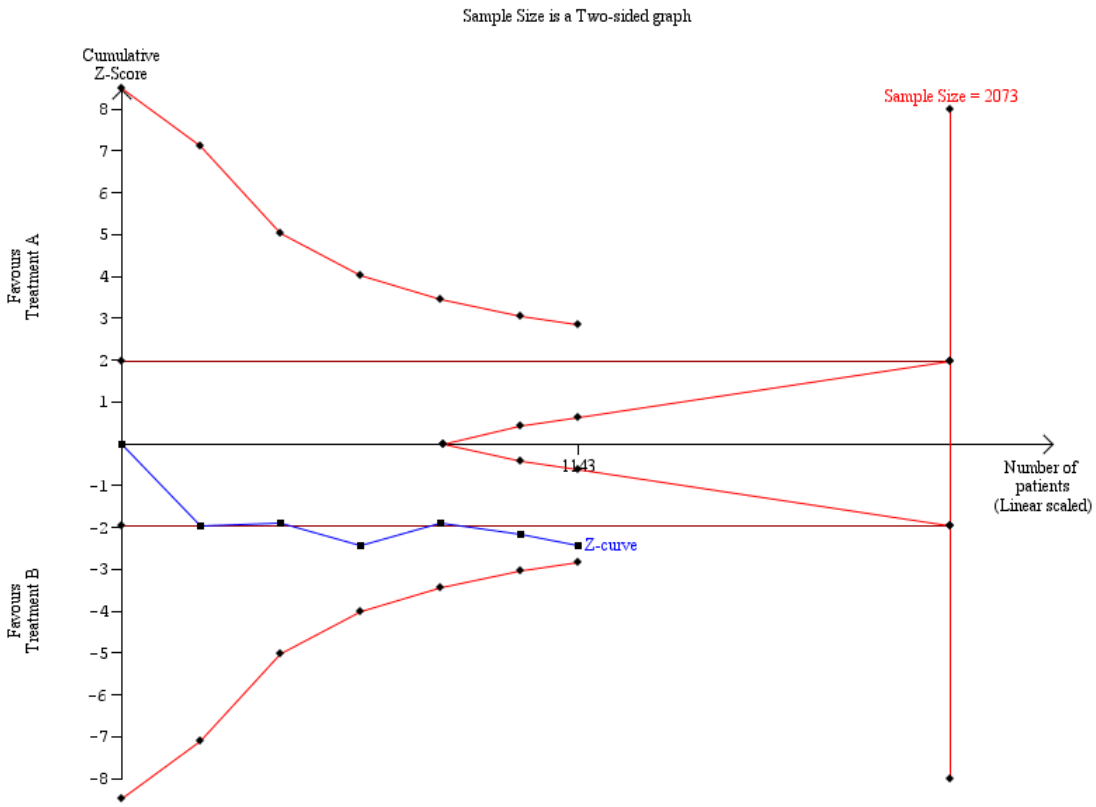

Figure 13 Example TSA showing the cumulative Z-score not crossing the monitoring boundary, and information size has not been reached.

3.5.5 Limitations of Trial Sequential Analysis

In the previous section, it has been explained how TSA may overcome the risks of type I and II errors when conducting meta-analyses. In recent years, TSA has been increasingly utilised by authors, the Cochrane

Collaboration and other evidence synthesis groups (108). However, TSA can be difficult to perform, may be misused and has its limitations (104, 113).

A criticism of TSA is that, if a TSA is designed and conducted following data collection, the analysis becomes data driven and may not be thorough enough to address a predefined alternative hypothesis (113). However, Wetterslev *et al.* (85) argued that many meta-analyses follow data-driven hypotheses and analyses. Therefore it is recommended that for each TSA, a protocol should be registered which describes the anticipated intervention effect, anticipated trial heterogeneity, and the anticipated outcome event rate in the comparator group prior to conducting the TSA (85). Alternatively, authors should make it explicit they are conducting a post-hoc analysis, and should do a sensitivity analysis around the values chosen to inform the TSA.

Recently, the Cochrane Collaboration evaluated and updated their guidance on using sequential approaches in meta-analysis in their systematic reviews (49, 50). The authors of the Cochrane Collaboration Handbook for Systematic Reviews of Interventions concluded that sequential methods should not be used in primary analyses or to draw conclusions, but could be used as secondary analyses in systematic reviews if they are prospectively planned and the assumptions underlying the design are clearly justified (50). In their guidance, they recommend that authors' interpretations of evidence should be based on estimated magnitude of intervention effect and its uncertainty rather than drawing binary conclusions from interpretations of the P-value from the TSA, and decisions should not be influenced by plans for future updates of meta-analyses (50). In the future there may be scope to use TSA in conjunction with the GRADE approach used in Cochrane reviews, to assess the certainty of the body of evidence relating to the outcomes. For example, if a TSA shows that more information is required,

this could be used as a reason to downgrade the certainty of an interventions effects.

Higgins *et al.* (113) questioned the analogy used in TSA between stopping trials based on interim analyses and 'stopping' further meta-analyses. In TSA, if there is sufficient or insufficient evidence to reject or not reject the null hypothesis, it is concluded that more studies are needed (113). If the null hypothesis is not rejected (the TSA result crosses the futility boundary, based on pre-specified power and minimum clinically important effect size) or rejected (the TSA result crosses boundaries that represent a harmful or beneficial effect), the research question has been answered, and no more studies are required (113). If these ideas are applied to a single trial, this can lead to the continuation or stopping of the trial. Higgins *et al.* (113) argues that this same notion cannot be applied to TSA since meta-analysts are not able to make these decisions about future studies, but should make recommendations instead. If new, high quality studies are already underway when the decision is made to stop further analyses, these would need to be included in updates, and should not be ignored.

It is argued that TSA relies too heavily on the result of the statistical significance test (P-value) rather than the 95% confidence intervals (114). In TSA, confidence intervals can be adjusted for the incomplete meta-analysis information size and for multiple significance testing (104). It has been suggested that the traditional 95% confidence intervals are sufficient enough to measure whether or not an intervention works (114), however these intervals exclusively relate to the null hypothesis and not to an alternative hypothesis relating to the type I error risk (115). If the unadjusted confidence intervals are used when the information size has not been reached, this can lead to false assertions of statistically significant events (104). Therefore, the traditional unadjusted 95% confidence interval

is only sufficient for statistical significance when the required information size has been achieved (104).

TSA has also been scrutinised for being too conservative as TSA users may decide to use a conservative *a priori* intervention effect and the total variance in the meta-analysis to calculate the required information size (104). Although using an *a priori* intervention effect does not consider the intervention effect from the collected data, doing so may lead to a greater required information size (116). Furthermore, although using the total variance for the calculation of information size is seen as the worst-case scenario of risk of random error, it is unknown whether this variation is produced by systematic differences or by random variations (104). As it cannot be deduced where the variation arises from, it must be assumed that all of the variance arises from chance (117).

Kulinskaya and Wood (118) have argued that in an underpowered meta-analysis, not only is it necessary to assess the gap from the accrued information size to the required information size (i.e. the number of additional participants you need to randomise), but also the number of studies that should be conducted to achieve the required information size (118). Using multiple studies to reach the required information size may be beneficial in meta-analyses where heterogeneity occurs (118) since smaller studies are more likely to have more imprecise estimates of intervention effects; hence contribute to the precision of the estimate of the between-study heterogeneity. However, setting up more than one study can be more expensive and this may not be realistic in practice.

3.6 Conclusion

TSA overcomes the issues of multiple testing resulting from updating a meta-analysis by providing corrected results using monitoring boundaries

and a required information size. TSA has the added advantage over standard meta-analysis methods, which allows the reader to assess whether there is sufficient evidence to conclude a clinically important treatment effect, no evidence of an effect, or lack of evidence. By giving an approximation for information size based on a minimum clinically important treatment effect, future studies can be better informed regarding sample size estimations. Furthermore, if information size has been surpassed, this can prevent further resources being wasted on more studies. In the following chapter an alternative use for TSA is presented, where it is used to estimate the sample size for a study based on results from feasibility and pilot trials.

Chapter 4: Using Trial Sequential Analysis for estimating the sample sizes of further trials

4.1 Introduction

The arguments presented in this chapter have been submitted in a manuscript to *BMC Medical Research Methodology*, and a pre-print has been published on *Research Square* (**Appendix B**) (119). Journal editors have requested a revised version which is currently under review.

Demonstrating that health interventions work requires substantial resources. Often feasibility and pilot randomised clinical trials are conducted before larger-scale randomised controlled trials (RCTs) are designed to determine benefits and harms (120-122). Feasibility trials are used to ascertain information such as intervention acceptability, feasibility of intervention delivery, and recruitment likelihood to help design more decisive RCTs (120). A pilot trial is a smaller version of a large-scale RCT, and is used to test whether the main components of the trial, such as recruitment, randomisation, treatment, and follow-up assessments can all work together (120). Moreover, their data can be used to inform sample sizes for large-scale RCTs (121, 122).

Chapter 3 discusses how TSA is a methodology that can be used in meta-analyses to control for random errors, and to assess whether further studies need to be conducted (123). In a novel approach, here we employ TSA and combine data from feasibility and pilot RCTs testing a text message-based smoking cessation intervention for pregnant women ('MiQuit') (124, 125) to estimate the sample size that one or more future RCTs would need to recruit, to provide a more decisive answer regarding intervention benefit.

4.2 Aims

The aim of this chapter is to demonstrate an alternative use for TSA by calculating the sample size required for an RCT of MiQuit, using results from feasibility and pilot studies.

4.3 Objectives

The study aim was investigated through the following objectives:

I. To use parameters from feasibility and pilot trials of MiQuit to perform TSA.
II. To use TSA to calculate the required information size of one or more trials of MiQuit.

4.4 Methods

As presented in **Section 3.5.1**, TSA can inform how much more information is required to yield a firm conclusion regarding the effect of the intervention versus its comparator – the distance between the accrued information and the required information.

In TSA, trials are chronologically ordered, and interim analyses are conducted as each trial is added. In a TSA where the 'required information size' has not been reached, the threshold for statistical significance is inflated to account for sparse data and multiple testing of the interim analyses using monitoring boundaries; thus, the 95% confidence interval is not providing coverage of the real uncertainty and the cut-off for determining statistical significance is below the usual nominal figure of 0.05 (104).

In the worked examples below, we show how TSA methods can be used to estimate the sample size required for one or more new trials to add further data to a meta-analysis to provide more firm evidence for an intervention either having or not having the postulated minimally clinically significant effect.

4.5 Results

In this section, we provide an example of how TSA successfully used data from feasibility and pilot RCTs that tested MiQuit, a text-message, self-help smoking cessation intervention for pregnant women, to justify research funds to undertake a third RCT.

4.5.1 Previous MiQuit trials

Smoking during pregnancy increases the risk of miscarriage, stillbirth, low birth-weight, premature birth, perinatal morbidity and mortality, sudden infant death, as well as adverse infant behavioural outcomes (126, 127). Pregnancy is a life event which motivates cessation attempts amongst smokers and over 50% of pregnant women who smoker attempt to quit during this time (128), consequently pregnancy is an opportune moment to offer smoking cessation support. Text message, self-help support, smoking cessation programmes developed for non-pregnant smokers are effective, but such programmes are inappropriate for use during pregnancy (129-131). To address the lack of acceptable self-help, support cessation programmes for pregnant smokers in the UK, MiQuit was developed (124). MiQuit delivers individually-tailored text messages to pregnant smokers, with the aim of encouraging them to stop smoking (124). Further details on MiQuit can be found elsewhere (124).

A MiQuit feasibility RCT was conducted, including 207 women. Biochemically-validated, 7-day point prevalence cessation at 12 weeks post randomisation (~6 months gestation) was 12.5% in the MiQuit group, compared with 7.8% in the control group (odds ratio (OR) 1.68, 95% confidence interval (CI) 0.90 to 3.16) (124). Although the trial was relatively small in sample size and the cessation period brief, the trial provided an estimate suggesting that MiQuit could have a positive impact in addition to routine care.

Next, a pilot RCT was conducted to investigate the feasibility of undertaking a fully-powered multi-centre RCT in UK National Health Service (NHS) settings (125). The pilot MiQuit RCT recruited 407 pregnant smokers and the prolonged abstinence rate from smoking, validated in late pregnancy was 5.4% in the MiQuit group versus 2.0% in the control group (OR 2.70, 95% CI 0.93 to 9.35) (125). This trial also suggested a beneficial effect of MiQuit.

As MiQuit is a cheap intervention and can be disseminated widely, we anticipated that even a 1% to 2% absolute effect on smoking cessation in pregnancy could be clinically important and cost effective (125). The results from the feasibility and pilot trials suggested that an impact of this size was attainable; however, an adequately powered RCT would still be needed to determine whether MiQuit is effective and guide future routine clinical practise.

4.5.2 Conventional meta-analysis

The conventional way to determine if an intervention is effective or not is to use the naïve alpha of 5% and the naïve 95% confidence interval (74). Since both the feasibility and pilot trials used virtually the same design as that which would be used in any new RCT, they can be considered as pilots and

it would be appropriate to meta-analyse these trials' findings together. Using a random-effects model, a traditional meta-analysis of pilot and feasibility studies' data found, that women randomised to MiQuit were more than twice as likely to be abstinent in their pregnancy (pooled OR 2.26, 95% CI 1.04 to 4.93; I^2=0%, p=0.041). Although, this result can be interpreted to be significant according to conventional assessment (P<0.05), it should be interpreted with caution because, as described above, findings from meta-analyses based on only two small RCTs can produce spurious findings due to type I error (86, 95, 132).

In the next sections, we use conventional sample size estimation methods to estimate the sample size for an RCT which, on its own would have enough power to show whether MiQuit might be effective, using a plausible treatment effect estimate derived from the conventional meta-analysis above. We also calculate a second sample size estimate for one or more further RCTs, which when pooled with data from feasibility and pilot trials using TSA methods, would be similarly decisive.

4.5.3 Conventional sample size estimation

As the pilot trial (125) was considered at lower risk of bias compared to the feasibility trial (124), a traditional sample size calculation using smoking cessation rate estimates derived from the pilot trial suggests a new trial would require a total sample size of 1292 participants. This estimate has 90% power (10% type II error) and 5% significance (2-sided test; type I error) to detect a 3.4% absolute difference in prolonged abstinence from smoking from 4 weeks after enrolment until 36 weeks' gestation between the MiQuit and control groups (5.4% versus 2.0%) (125).

4.5.4 Trial Sequential Analysis

Figure 14.I illustrates a TSA incorporating findings from the MiQuit feasibility (A) (124) and pilot (B) (125) trials. In this TSA output, the x-axis represents the number of participants and marked on this are the numbers of participants recruited to each trial. The y-axis represents the Z-score, where a positive Z-score favours the MiQuit intervention and a negative Z-score favours the control.

The Z-score is the test that helps you decide whether to reject or not reject the null hypothesis. Very high positive or very low negative Z-scores are associated with very small P-values. The critical Z-score values when using a 95% confidence level which are known as the 'conventional test boundaries', are -1.96 and +1.96 and these relate to a two-sided P-value of 0.05. If the Z-score is between -1.96 and +1.96, the P-value will be larger than 0.05, and the null hypothesis of no difference between intervention groups is not rejected. The Z-curve represents the cumulative Z-score as each RCT is added to the analysis. In **Figure 14.I**, when trial B is added to the analysis, the Z-curve crosses the conventional test boundary (p=0.05). This is consistent with the results from the conventional meta-analysis for MiQuit, where we found P=0.041.

The required information size is represented by the vertical red line in **Figure 14**. The required information size was estimated using the same parameters as used for the conventional sample size estimation above (90% power, 5% significance, to detect a 3.4% absolute difference) (125); although this estimate could take into account observed heterogeneity, there was none in this meta-analysis due to the similarity of the intervention and methodology used within the trials (I^2 = 0% and D^2 = 0). Consequently, the estimated required information size of 1296 participants is only slightly

different to that using conventional sample size estimation due to rounding errors; however, the estimates would be larger if heterogeneity were present.

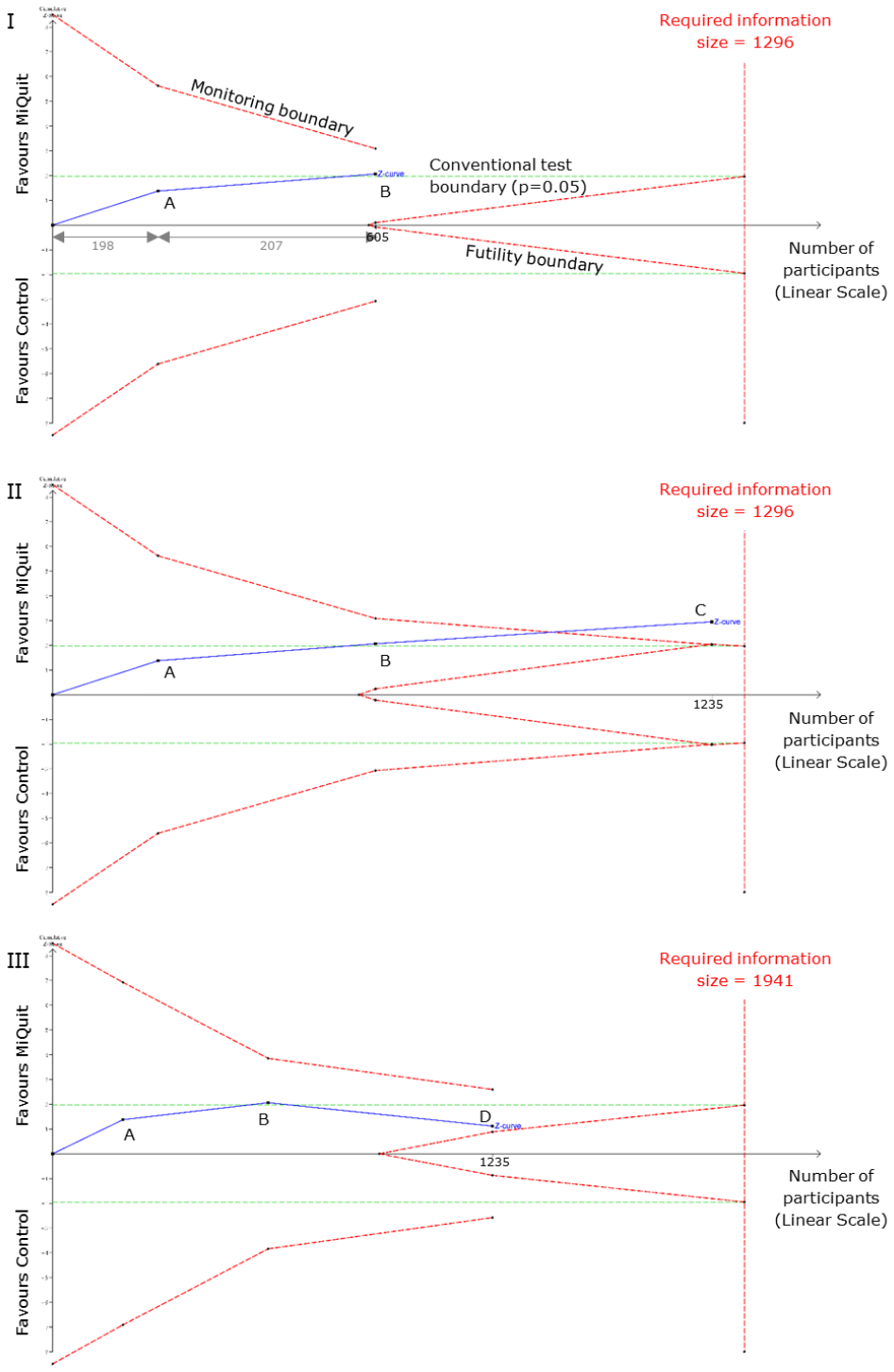

Figure 14 Trial Sequential Analysis output of both MiQuit trials using; 90% power, 5% significance, to detect a 3.4% absolute difference. Points A and B on the Z-curve represent each trial added to the trial sequential analysis. A – Feasibility trial n=198 (124); B – Pilot trial n=407 (125). **Figure 14.II.** Point C represents a theoretical trial with a sample size of 630 women, where an absolute difference of 3.17% was observed, in favour of the MiQuit group, between the control and intervention groups. **Figure 14.III.** Point

D represents a theoretical trial with a sample size of 630 women, with an absolute difference of -0.63% in favour of the control group.

As the cumulative Z-curve does not cross the upper trial sequential monitoring boundary which indicates MiQuit being effective, this TSA shows that further information is required before any firm conclusion can be reached about the efficacy of the MiQuit intervention. Although the conventional meta-analysis suggested, with borderline significance, that pregnant women randomised to MiQuit were more than twice as likely to be abstinent from smoking in late pregnancy, the TSA indicates that this finding is not sufficiently robust. The TSA-adjusted 95% confidence intervals for cessation using MiQuit (pooled OR 2.26, TSA-adjusted 95% CI 0.66 to 7.70), are much wider than those of the conventional meta-analysis (pooled OR 2.26, unadjusted 95% CI 1.04 to 4.93).

Without TSA having been undertaken, an interpretation of the conventional meta-analysis would have been that MiQuit is effective. However, TSA indicates that one cannot be secure in this interpretation and further trial data should be collected to eliminate the possibility that this is a false positive result, which can occur early in intervention evaluation, particularly when small trials are undertaken.

4.5.5 Calculating sample size for a third MiQuit RCT

TSA has demonstrated that further RCT data are required before a firm conclusion about the efficacy of the MiQuit intervention can be determined. As the initial two trials were sufficiently similar to be combined in the TSA, we will now demonstrate how TSA methods can be used to estimate the sample size for (a) further trial(s) – data from which, when combined with the previous two trials in the TSA, would be expected to provide more conclusive findings regarding the efficacy of the MiQuit intervention. We will also demonstrate how exemplar theoretical findings from future trials which

are both in favour and against MiQuit having a positive effect would impact the TSA result.

4.5.5.1 Trial Sequential Analysis sample size estimation

Estimates derived from the TSA found the required information size as 1296 participants. From the feasibility and pilot studies, 605 women have already been recruited and randomised; therefore, the required sample size for further RCTs can be estimated as the difference between the required information size minus the number of women already recruited into the previous trials; thus a sample size of 691 women (346 per intervention group) would be needed, assuming a 1:1 ratio.

Figure 14.II shows the TSA output after adding a theoretical third trial (C) with a sample size of 630 women (315 per trial group), where an absolute difference of 3.17% was observed in favour of the MiQuit group versus the control group. The TSA shows the cumulative Z-curve line crossing the upper trial sequential monitoring boundary which indicates MiQuit being effective. As the trial sequential monitoring boundary has been crossed, the TSA Z-curve does not need to reach the required information size of 1296. In the present scenario, we can firmly conclude that MiQuit is effective for smoking cessation compared to control (provided that all trials are valid and not influenced by systematic errors (bias) or other errors).

When a theoretical third trial (D) with a negative outcome is included in the TSA (**Figure 14.III**), we observe a different conclusion. Here, the third trial D with a sample size 630 was intentionally given a negative outcome (absolute difference of -0.63% in favour of control). Here we observe the Z-curve drop below the conventional test boundary, and in a meta-analysis we would have concluded that MiQuit was not effective. However, in the TSA, the futility boundary is not crossed, so we are unable to decisively say

that MiQuit is not as effective as control for smoking cessation. Due to the diversity, the required information size has increased to 1941, meaning future trials will need a further 706 participants.

4.5.5.2 A conservative approach to sample size estimation

In the above example, the required information size was derived using the smoking cessation rate from the pilot trial (125). Therefore, it can be postulated that data from the pilot trial should not be included in subsequent TSA. Consequently, consistent with this one could exclude the data from the pilot trial from the TSA and re-estimate the total number required (**Figure 15.I**). Using this approach, to provide a conclusive result, either a single trial of 1098 participants (549 per intervention group, assuming a 1:1 ratio) or multiple trials cumulating to a total of 1098 participants, would be needed. This figure, although conservative, is still less than the estimate from the conventional sample size calculation.

Figure 15.II and **Figure 15.III** also show the TSA outputs if theoretical trials C and D were included in the TSA. In both situations further information is needed, despite the Z-curve coming close to the upper trial sequential monitoring boundary in **Figure 15.II** and the futility boundary in **Figure 15.III**.

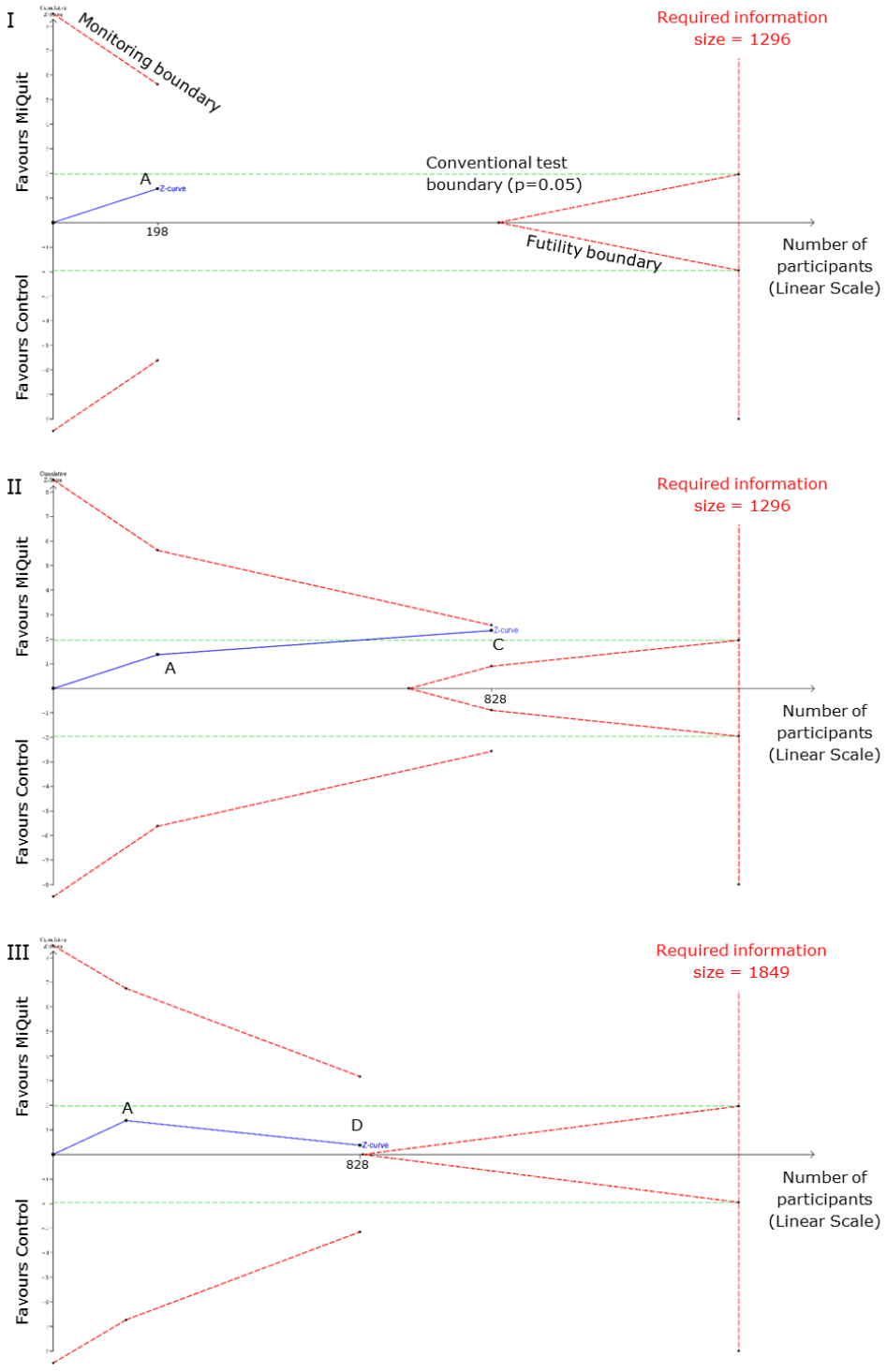

Figure 15.I Trial Sequential Analysis output of the MiQuit feasibility trial with the pilot trial removed, using; 90% power, 5% significance, to detect a 3.4% absolute difference. Point A on the Z-curve represents the feasibility trial. **Figure 15.II**. Point C represents a theoretical trial with a sample size of 630 women, where an absolute difference of 3.17% was observed, in favour of the MiQuit group, between the control and intervention groups. **Figure 15.III**. Point D represents a theoretical trial with a sample size of

630 women, with an absolute difference of -0.63% in favour of the control group.

4.5.6 Sensitivity analysis

The modelled scenario, in which there is no heterogeneity between trials in a meta-analysis is rare; in most situations where the described approach is used, some heterogeneity between studies might be expected. TSA provides 95% confidence intervals for heterogeneity (I^2) within meta-analyses. One way to fully allow for heterogeneity is to perform a sensitivity analysis using the upper boundary for heterogeneity. This would increase the required information size. In our example, the program could not calculate the 95% confidence interval surrounding the I-square of 0% as there were less than three included studies. In this case it is possible to input an estimate for heterogeneity into the TSA software.

4.6 Discussion

The chapter demonstrates how TSA can be used to determine the required sample size for one or more additional RCTs to make the findings from a meta-analysis more conclusive. This sample size would be considered underpowered in comparison to a traditional single RCT sample size calculation. However, by using TSA in such a way, future trials could be planned using significantly fewer resources and with less cost than trials planned using traditional sample size calculations.

In the worked example, data from the pilot trial was used in the TSA to estimate the required information size. Ignoring that the same data is being used twice (for the estimation and for the meta-analysis) could mean that the estimate generated is not sufficiently conservative. Thus, we present a modification which attempts to overcome this issue. This approach increases

the difference between required information size minus the accrued information by the sample size of the trial used in the estimation.

It is important to note that in the example, the meta-analysis of the existing two MiQuit trials quantified heterogeneity as 0%, thereby indicating that none of the variation in the meta-analysis was due to heterogeneity. However, it is unlikely that this will be the case for meta-analyses of other interventions aimed at changing addictive behaviours (133, 134); therefore, TSA methods have been developed to account for this (132). In TSA, estimated information size and monitoring boundaries vary with the level of heterogeneity in the meta-analysis, where the greater the level of heterogeneity, the larger the sample size needed for firm conclusions to be reached.

In the example presented, odds ratios were also used instead of risk ratios, as the feasibility study was powered using an odds ratio from a meta-analysis investigating mobile phone interventions for smoking cessation in the general population (124). Moreover, the quit rates are relatively low, so there is very little difference between the odds ratio and relative risk. In other TSAs, it may be advisable to use risk ratios instead of odds ratios, to avoid overestimating the intervention effect. Additionally, it may be inappropriate to use the odds ratio used to power the feasibility trial to estimate sample sizes for future MiQuit trials since data now exists from the feasibility and pilot trials. In our example, the stipulated intervention effect was derived from the pilot trial ('internal data'), and it could be argued that such adaptive data should not be used in meta-analysis (135).

In **Chapter 3.5.5** it was discussed that using multiple trials to reach the required information size may be beneficial in meta-analyses where heterogeneity occurs (118). Smaller trials tend to have more imprecise

estimates of intervention effects; hence contribute to the estimation of heterogeneity in the meta-analysis. However, setting up more than one trial can be more expensive, and this may not be realistic in practice.

In **Chapter 3.5.5** it was also discussed how authors of the Cochrane Collaboration Handbook for Systematic Reviews of Interventions had reservations regarding using TSA to draw conclusions regarding the effectiveness of an intervention (136, 137). However, these criticisms of sequential approaches in meta-analyses apply to the traditional use of TSA, whereas this chapter demonstrates an alternative use of the method. Furthermore, a further limitation the authors argue is that a meta-analyst does not have any control over the future designing of trials that would be eligible for inclusion in the meta-analysis (66), thereby making it impossible to construct a set of stopping rules (66). However, in our example, the opposite is the case, where both the feasibility and pilot trials were conducted by the same group of investigators, and any future trials would have a consideration for the desired properties of a stopping rule.

Finally, the authors of the Cochrane Collaboration Handbook for Systematic Reviews of Interventions also highlight that there are methodological limitations to sequential methods when heterogeneity is present (137). In our example described in this chapter, heterogeneity was not present and therefore these limitations are not relevant. However, we discuss how the presence of heterogeneity could be explored in TSA by performing sensitivity analyses.

4.7 Conclusions

In conclusion, TSA is a method that can utilise data from feasibility and pilot trials as well as other trials, in order to estimate a sample size for one or more future RCTs, to provide an adequately powered conclusion regarding

an intervention’s benefits and harms. This simple use of expensively collected trial data could be usefully exploited by researchers evaluating other interventions and could result in cost saving as fewer participants would need to be recruited than if a conventional sample size estimate is carried out.

Chapter 5: A systematic review of nicotine replacement therapy for promoting smoking cessation during pregnancy

5.1 Introduction

The work presented in this chapter is an update to an existing Cochrane review (59), and has been published as part of wider systematic reviews of 'Pharmacological interventions for promoting smoking cessation during pregnancy' published in the *Cochrane Library* (**Appendix B**); and 'Fetal safety of nicotine replacement therapy in pregnancy: systematic review and meta-analysis' published in *Addiction* (**Appendix B**).

Chapter 1 outlined the epidemiology of smoking in pregnancy and described the risks associated with smoking in pregnancy. There was also a description of how NRT is effective for smoking cessation outside of pregnancy, but its efficacy in pregnancy is unclear. **Chapter 1** also discussed some of the safety concerns surrounding NRT use in pregnancy. In this chapter a systematic review is performed to assess the efficacy of NRT for smoking cessation in pregnancy, and how safe NRT is when used in pregnancy.

5.1.1 Why it is important to do this review

Guidelines from many countries recommend that NRT be offered for smoking cessation in pregnancy to heavy smokers who have been unable to quit smoking using behavioural or psychosocial methods (45, 138-141). In most high-income countries (e.g. Canada, the USA, Australia, New Zealand), guidelines recommend that pregnant women be offered intermittent NRT-delivery formulations (e.g. gum, lozenges, spray - classified as category C drugs in pregnancy), rather than continuous ones (e.g. patches - classified as category D) (142). The theoretical rationale for this is that the overall dose of nicotine delivered by intermittent formulations may be lower than that delivered by continuous ones (140), and that the peaks in blood nicotine concentrations are more extreme, mimicking the

action of smoking. However, some experts recommend patches, as the lower peak nicotine levels associated with these may induce fewer adverse effects, such as throat irritation (45, 140).

Consensus-based recommendations about using NRT for smoking cessation in pregnancy are underpinned by a belief that medicinal NRT is safer than smoking (143). However, to date, individual trials have had inconsistent findings (55, 58), and there is no conclusive evidence that NRT is either effective or safe in pregnancy (144). There are also reports of low adherence to NRT regimens, which could reduce efficacy and suggests that the acceptability of NRT use in pregnancy may be limited (52, 145). Furthermore, it is unclear whether efficacy or safety is improved with intermittent NRT administration (fast-acting NRT products) or with continuous administration using nicotine patches.

Given that NRT appears to be widely accepted for cautious use in pregnancy, a systematic review investigating the efficacy and safety of this clinical practice was warranted. An up-to-date, robust synthesis of research evidence on the use of NRT for cessation in pregnancy will help advance clinical practice in an area of substantial clinical need.

5.2 Objectives

To determine the efficacy and safety of NRT used during pregnancy for smoking cessation in later pregnancy and after childbirth, and to determine adherence to NRT for smoking cessation during pregnancy.

5.3 Methods

5.3.1 Criteria for considering studies for this review

5.3.1.1 Types of studies

Parallel- or cluster-randomised controlled trials (RCTs) were eligible for inclusion. Quasi-randomised, cross-over, and within-participant designs were not eligible for inclusion due to the potential biases inherent in these designs.

5.3.1.2 Types of participants

Women who were pregnant and who also smoked tobacco at study baseline.

5.3.1.3 Types of interventions

Comparisons of any type of NRT (including chewing gum, transdermal patches, nasal and oral spray, inhalators and tablets or lozenges) versus placebo or no NRT control.

Trials could provide behavioural support to participants, however the support provided had to be very similar (ideally identical) across the active NRT and comparator trial arms. Behavioural support is effective for smoking cessation in pregnancy (40), and differences in its provision would be expected to affect cessation and birth outcomes, potentially rendering findings difficult to interpret.

5.3.1.4 Types of outcome measures

5.3.1.4.1 Primary outcomes

Self-reported abstinence from smoking at the latest time point in pregnancy at which this was measured and, where available, validated biochemically using measures such as exhaled carbon monoxide, saliva cotinine, or, in

those who are not smoking but using nicotine, anabasine. When validated abstinence data were available, these were preferred to self-report. Where this information was available, we also used prolonged or continuous abstinence measures, timed from a quit date set in early pregnancy and which allowed temporary lapses to smoking as per the Russell Standard criteria for outcome measurement in cessation studies (146). However, point prevalence abstinence measures were substituted for these as required.

5.3.1.4.2 Secondary outcomes

1) Abstinence from smoking after childbirth (with abstinence defined as detailed above)
2) Safety
    a) Miscarriage/spontaneous abortion
    b) Stillbirth
    c) Mean unadjusted birthweight
    d) Low birthweight (less than 2500 g)
    e) Preterm birth (less than 37 weeks' gestation)
    f) Neonatal intensive care unit admissions
    g) Neonatal death
    h) Caesarean section
    i) Congenital anomaly
    j) Maternal hypertension
    k) Infant respiratory symptoms
    l) Infant development
3) NRT adherence
4) Non-serious adverse effects (serious adverse event data contributed to safety outcomes, as described above)
5) Any reported long-term effects of NRT on safety

We did not carry out a specific literature search for outcomes 3 to 5, but, if reported, these data were extracted from the included studies and described qualitatively.

5.3.2 Search methods for identification of studies

5.3.2.1 Electronic searches

We searched the Cochrane Pregnancy and Childbirth Group's Trials Register by contacting their Information Specialist, who ran the search on 20 May 2019.

The Cochrane Pregnancy and Childbirth Group's Trials Register is maintained by their Information Specialist and contains trials identified from:

1) monthly searches of the Cochrane Central Register of Controlled Trials (CENTRAL);
2) weekly searches of MEDLINE (Ovid);
3) weekly searches of Embase (Ovid);
4) monthly searches of CINAHL (EBSCO) (Cumulative Index to Nursing and Allied Health Literature);
5) hand-searches of 30 journals and the proceedings of major conferences;
6) weekly current awareness alerts for a further 44 journals plus monthly BioMed Central email alerts;
7) scoping searches of the US National Institutes of Health Ongoing Trials Register ClinicalTrials.gov (clinicaltrials.gov/) and the World Health Organization International Clinical Trials Registry Platform (apps.who.int/trialsearch/) for unpublished, planned, and ongoing trial reports.

Based on the intervention described, each trial report is assigned a number that corresponds to a specific Pregnancy and Childbirth review topic (or topics) and is then added to the Register. The Information Specialist searches the Register for each review using this topic number rather than keywords. This results in a more specific search set that has been fully accounted for in the relevant review sections (Included, Excluded, Awaiting Classification, or Ongoing).

Details of the search strategies for CENTRAL, MEDLINE, Embase, and CINAHL; the list of hand searched journals and conference proceedings; and the list of journals reviewed via the current awareness service can be found in the 'PCG Trials Register' section of the Cochrane Pregnancy and Childbirth Group's website.

5.3.2.2 Searching other resources

We checked relevant cited studies whilst reviewing the trial reports identified by the electronic searches, as well as reference lists from any directly relevant reviews identified. We also searched the following trials registers on 20 May 2019: US National Institutes of Health Ongoing Trials Register ClinicalTrials.gov (clinicaltrials.gov/) and the World Health Organization International Clinical Trials Registry Platform (apps.who.int/trialsearch/), and OpenGrey, "System for Information on Grey Literature in Europe" (www.opengrey.eu/).

We did not apply any language or date restrictions and included studies regardless of the publication type (e.g. conference abstract, trial registry entry, journal article).

5.3.3 Data collection and analysis

For this update, the following methods were used to assess the newly identified studies resulting from the latest search.

5.3.3.1 Selection of studies

Two review authors (RC and TC) independently inspected the search results, making separate lists of titles and abstracts that were potentially suitable for inclusion. We then retrieved the full texts of reports deemed potentially relevant, and two review authors (RC and TC) independently assessed these for inclusion in the review. At both stages disagreements were resolved by discussion without the need to involve a third review author.

5.3.3.2 Data extraction and management

We designed a data extraction form based on that used by Lumley *et al*., 2009 (147), which two review authors (RC and TC) used to independently extract data from eligible studies. Extracted data were compared, with any discrepancies being resolved through discussion. RC entered data into Review Manager 5 software (148), double checking this for accuracy.

When information regarding any of the above was unclear, we contacted authors of the reports to provide further details.

We recorded the following information, where available, in a 'characteristics of included studies' tables (**Appendix A**).

1) Methods: study design.
2) Participants: number of participants, inclusion criteria, and any relevant exclusion criteria.

3) Interventions: description of intervention and control (treatment, dosage, regimen, behavioural support, duration of intervention), information regarding dose matching if relevant.
4) Outcomes: primary outcomes, time points reported, biochemical validation, and definitions of abstinence.
5) Notes: we recorded dates of the trial, trial funding, and declarations of interest of trial authors where reported.

We created additional tables for details of twin births and fetal loss in pregnancy and for extracted adherence data. Adherence data can be found in **Appendix A**.

5.3.3.3 Assessment of risk of bias in included studies

RC and TC independently assessed risk of bias for all studies which they had not authored (the one study led by TC was assessed by CC and JLB), using criteria adapted from those in the *Cochrane Handbook for Systematic Reviews of Interventions* (137). Any disagreements were resolved by discussion with a third review author (JLB).

We assessed the following 'Risk of bias' domains for all included studies.

5.3.3.3.1 Random sequence generation (checking for possible selection bias)

We determined whether the method used to generate the allocation sequence was sufficiently described to permit an assessment of whether it should produce comparable groups.

We assessed the method as:

- low risk of bias (any truly random process, e.g. random number table; computer random number generator);

- high risk of bias (any non-random process, e.g. odd or even date of birth; hospital or clinic record number); or
- unclear risk of bias.

5.3.3.3.2 Allocation concealment (checking for possible selection bias)

We determined the method used to conceal the allocation sequence and whether intervention allocation could have been foreseen in advance of or during recruitment, or changed after assignment.

We assessed the methods as:

- low risk of bias (e.g. telephone or central randomisation; consecutively numbered, sealed, opaque envelopes);
- high risk of bias (open random allocation; unsealed or non-opaque envelopes, alternation; date of birth); or
- unclear risk of bias.

5.3.3.3.3 Blinding (checking for possible performance bias and detection bias)

In smoking cessation studies, bias can also occur at outcome ascertainment if trial participants report that they have stopped smoking when actually they have not. Generally, it is perceived that the broadly negative social view of smoking can result in self-perceived pressure on participants in smoking cessation studies to be seen as having successfully stopped smoking, and this may result in false reporting of abstinence from smoking at follow-up. Trialists attempt to minimise this bias (detection bias) through use of biochemical validation of self-reported smoking status data which is collected for trial outcomes.

We determined the methods used, if any, to blind study participants and personnel from knowledge of which intervention was received by the participant. In the previous version of this review, we categorised studies that used placebo as at low risk of bias and those that used a behavioural control only as at high risk of bias. Using this categorisation of bias, findings with respect to efficacy of NRT were different for placebo (low risk of bias) and non-placebo (high risk of bias) RCTs, so we have maintained the same classification for this update. In the 'Risk of bias' table we also note whether participants, personnel, and outcome assessors were blinded to outcome assessment and whether the abstinence outcome was biochemically validated. We used cut points derived by expert consensus: 8 parts per million where exhaled carbon monoxide was used for validation and 10 ng/mL for saliva cotinine.

5.3.3.3.4 Incomplete outcome data (checking for possible attrition bias through withdrawals, dropouts, protocol deviations)

We determined for the primary outcome (i.e. smoking cessation) the completeness of data including attrition and exclusions from the analysis and whether an intention-to-treat analysis (i.e. reporting trial arm cessation rates amongst all participants who were originally randomised to that arm) was reported. We assessed whether attrition and exclusions were reported, the numbers included in the analysis at each stage (compared with the total randomised participants), reasons for attrition or exclusion where reported, and whether missing data were balanced across groups or were related to outcomes.

5.3.3.3.5 Selective reporting bias

We determined the possibility of selective outcome reporting bias and assessed methods as:

- low risk of bias (where it was clear that all of the study's prespecified outcomes and all expected outcomes of interest to the review have been reported);
- high risk of bias (where a prespecified outcome is not reported and there is evidence that this is due to lack of effect or an effect deemed unfavourable); or
- unclear risk of bias (where not all of the study's prespecified outcomes have been reported; one or more reported primary outcomes were not prespecified; outcomes of interest are reported incompletely and so cannot be used; or the study fails to include results of a key outcome that would have been expected to have been reported, however there is no clear evidence that this is a source of bias).

5.3.3.3.6 Other risk of bias

We considered whether there were any other additional potential sources of bias in the study.

5.3.3.3.7 Overall risk of bias

Where a study was judged to be at low risk for all of the above domains, it was considered to be at overall low risk of bias; where at least one judgement of high risk of bias was made, the study was considered to be at overall high risk of bias; and where there was no judgement of high risk, but at least one judgement of unclear risk, the study was considered to be at overall unclear risk of bias.

5.3.3.4 Assessment of the certainty of the evidence using the GRADE approach

We used the GRADE approach to assess the certainty of the body of evidence relating to the following outcomes for each comparison (NRT

versus control) (149), as recommended in the *Cochrane Handbook for Systematic Reviews of Interventions* (137):

- smoking cessation at the latest point in pregnancy (primary outcome);
- mean birthweight (safety outcome). We chose mean birthweight because it can be used as a marker of multiple infant safety outcomes;
- miscarriage and spontaneous abortion (safety outcome). We chose this alongside mean birthweight because it is an important safety outcome that would not be reflected in the above mean birthweight outcome.

We used GRADEpro GDT to import data from Review Manager 5 in order to create a 'Summary of findings' table (**Table 1**) (148, 150). A summary of the intervention effect and a measure of certainty for the above outcomes was produced using the GRADE approach. The GRADE approach uses five considerations (study limitations, consistency of effect, imprecision, indirectness, and publication bias) to assess the certainty of the body of evidence for each outcome. The evidence can be downgraded from 'high certainty' by one level for serious (or two levels for very serious) limitations, depending on each of these considerations.

5.3.3.5 Measures of treatment effect

5.3.3.5.1 Dichotomous data

For dichotomous data (all outcomes except mean birthweight), including smoking cessation, we have presented results as summary risk ratios (RR) with 95% confidence intervals (CI). A RR > 1 for the smoking cessation outcomes indicates benefit of the intervention. For undesirable outcomes, such as preterm births, RR < 1 indicates benefit of the intervention.

5.3.3.5.2 Continuous data

For mean birthweight (continuous data), we have presented the mean difference (MD) between control and intervention groups with 95% CI.

5.3.3.6 Unit of analysis issues

5.3.3.6.1 Multiple pregnancies

The unit of analysis for smoking cessation was the trial participant, regardless of whether she had a singleton or multiple pregnancy. For all other outcomes, analyses were conducted amongst singleton births only; this approach was undertaken because adverse pregnancy events/outcomes, adverse infant birth outcomes, and poorer infant development are strongly associated with multiple pregnancy. Hence, analysing multiple and singleton pregnancies together for these outcomes could render review findings difficult to interpret. Outcome data from multiple births were insufficient for these to be analysed separately.

5.3.3.6.2 Cluster-randomised trials

This study design was eligible for inclusion, however no cluster-randomised trials were identified. If in future updates such trials are identified, we will include them in the analyses along with individually randomised trials. We will adjust their sample sizes or standard errors using the methods described in Sections 16.3.4 and 16.3.6 of the *Cochrane Handbook for Systematic Reviews of Interventions* (137), employing an estimate of the intracluster correlation co-efficient (ICC) derived from the trial (if possible), from a similar trial, or from a study of a similar population. If we use ICCs from other sources, we will report this and conduct sensitivity analyses to investigate the effect of variation in the ICC. If we identify both cluster-randomised trials and individually randomised trials, we will synthesise the

relevant information. We will consider it reasonable to combine the results from both if there is little heterogeneity between the study designs, and the interaction between the effect of intervention and the choice of randomisation unit is considered to be unlikely.

We will also acknowledge heterogeneity in the randomisation unit and perform a sensitivity analysis to investigate the effects of the randomisation unit.

5.3.3.7 Dealing with missing data

For the primary smoking abstinence outcome, we assumed any participants lost to follow-up were still smoking or had relapsed to smoking, using the Russell Standard criteria (146). At all outcome points, participants whose smoking status was unknown were assumed to be smoking.

We used the following denominators for other outcomes.

- For the pre-birth outcomes, miscarriage/spontaneous abortion and stillbirth, the denominator used was the number of women randomised with viable singleton pregnancies at the time of randomisation. Where terminations occurred after randomisation, terminated fetuses were excluded from the denominator if terminations were performed on a presumed viable fetus for non-medical reasons. Similarly, pregnancies that were documented as non-viable at the point of randomisation were also excluded from this denominator (e.g. missed abortion). Where terminations were undertaken for medical reasons and were judged incompatible with life, these cases were included in denominators and also within numerators; they were counted as miscarriages if performed before 24 weeks, and as stillbirths if conducted after this time point.

- For mean unadjusted birthweight (i.e. the only birth outcome measured on a continuous scale), the denominator used was the number of singleton births for which this outcome was recorded.
- For dichotomous birth outcomes (e.g. low birthweight, preterm birth, neonatal intensive care admissions, and neonatal death), the denominator used was the number of live births from singleton pregnancies.
- For infant outcomes, the number of live births was used.

For selected secondary outcomes and where appropriate and feasible, we conducted sensitivity analyses to investigate the impact of missing data on pooled treatment effect estimates.

For all outcomes, we carried out analyses, to the greatest degree possible, on an intention-to-treat basis (caveats outlined above); we attempted to include all participants randomised to each group in analyses, and all participants were analysed in the group to which they had been allocated regardless of whether or not they received the allocated intervention.

5.3.3.8 Assessment of heterogeneity

We assessed statistical heterogeneity in each meta-analysis visually by inspecting the overlap of 95% CIs for the individual studies on the forest plots. We quantified heterogeneity using the I^2 statistic (137). We regarded heterogeneity as substantial and hence worthy of further investigation (see Subgroup analysis and investigation of heterogeneity) if the I^2 was greater than 50%.

5.3.3.9 Assessment of reporting biases

As there were fewer than 10 studies in all meta-analyses, we did not draw funnel plots to assess the potential for reporting bias. If in future updates

of this review there are 10 or more studies, we will investigate reporting biases (such as publication bias) using funnel plots. We will assess funnel plot asymmetry visually if asymmetry is suggested by a visual assessment, and we will perform exploratory analyses to investigate it.

5.3.3.10 Data synthesis

We carried out statistical analysis using Review Manager 5 software (148). Following the standard methods of the Cochrane Tobacco Addiction Group for pharmacological interventions, we elected to use a fixed-effect model for meta-analyses of smoking abstinence data. For meta-analyses of safety and adverse events data, we used random-effects models, as effects are likely to vary across populations due to significant differences in baseline risk.

5.3.3.11 Subgroup analysis and investigation of heterogeneity

We performed an exploration of heterogeneity for primary and secondary outcomes where the I^2 was greater than 50%. Additionally, for smoking cessation outcomes, we performed subgroup analyses based on the following groups.

1) Placebo-controlled versus non-placebo-controlled RCTs
2) Studies using different types of NRT, both alone and in combination (i.e. fast-acting NRT and nicotine patch)
3) Low-dose NRT (< 10 mg/24 hours) versus high-dose NRT (> 10 mg/24 hours)

For secondary outcomes, where the I^2 was greater than 50% (indicating substantial heterogeneity), we also performed these subgroup analyses as an exploration of heterogeneity; however, they were not conducted routinely for all secondary outcomes due to too few studies included in the meta-analyses.

We assessed differences between subgroups statistically using subgroup interaction tests, and have presented the P values from these tests.

If in future updates of the review more than 10 studies are included in a meta-analysis, we may consider performing meta-regression to further explore reasons for heterogeneity or to analyse adherence data. A caveat to using this method for adherence data is that there is currently no standard method for reporting adherence; however, for meta-regression to be undertaken, studies must report adherence data similarly.

5.3.3.12 Sensitivity analysis

We planned two sensitivity analyses using smoking cessation outcomes, depending on the availability of data.

1) Excluding studies rated at high risk of bias overall.
2) Excluding any studies that reported substantially lower treatment adherence than others. As there is no consensus on what constitutes good or acceptable adherence to NRT in pregnancy, we anticipated defining 'low adherence' after consideration of adherence data reported within the included studies.

We were unable to carry out these analyses for the current review (explanations follow in the Results section); they will be undertaken in future review updates, data permitting.

5.4 Results

5.4.1 Results of the search

We carried out an updated search of the Cochrane Pregnancy and Childbirth Group's Trials Register on 20 May 2019 and identified 14 trial reports for potential inclusion. We also deemed a further study, which had recently

been published and so was not identified by searches, as potentially relevant (151). We identified a total of 15 trial reports for title and abstract screening, of which eight studies were clearly not RCTs and were excluded.

We obtained the full text of the seven remaining records for screening. We excluded one article (152), assessed four articles as ongoing studies (see below), and included one article in this update (151). Details of the flow of studies for this update are recorded in a PRISMA diagram in **Figure 16**. Eight trials included in previous versions of this review are also included in this update (51-58).

This updated review therefore includes a total of 9 trials (30 reports). It contains data from one additional trial published since the previous version (151), and involves a total of 2336 pregnant women who smoked at study baseline. We added two newly identified follow-up reports for each of two previously included trials, Coleman *et al*., 2012 (52) and Berlin *et al*., 2014 (51).

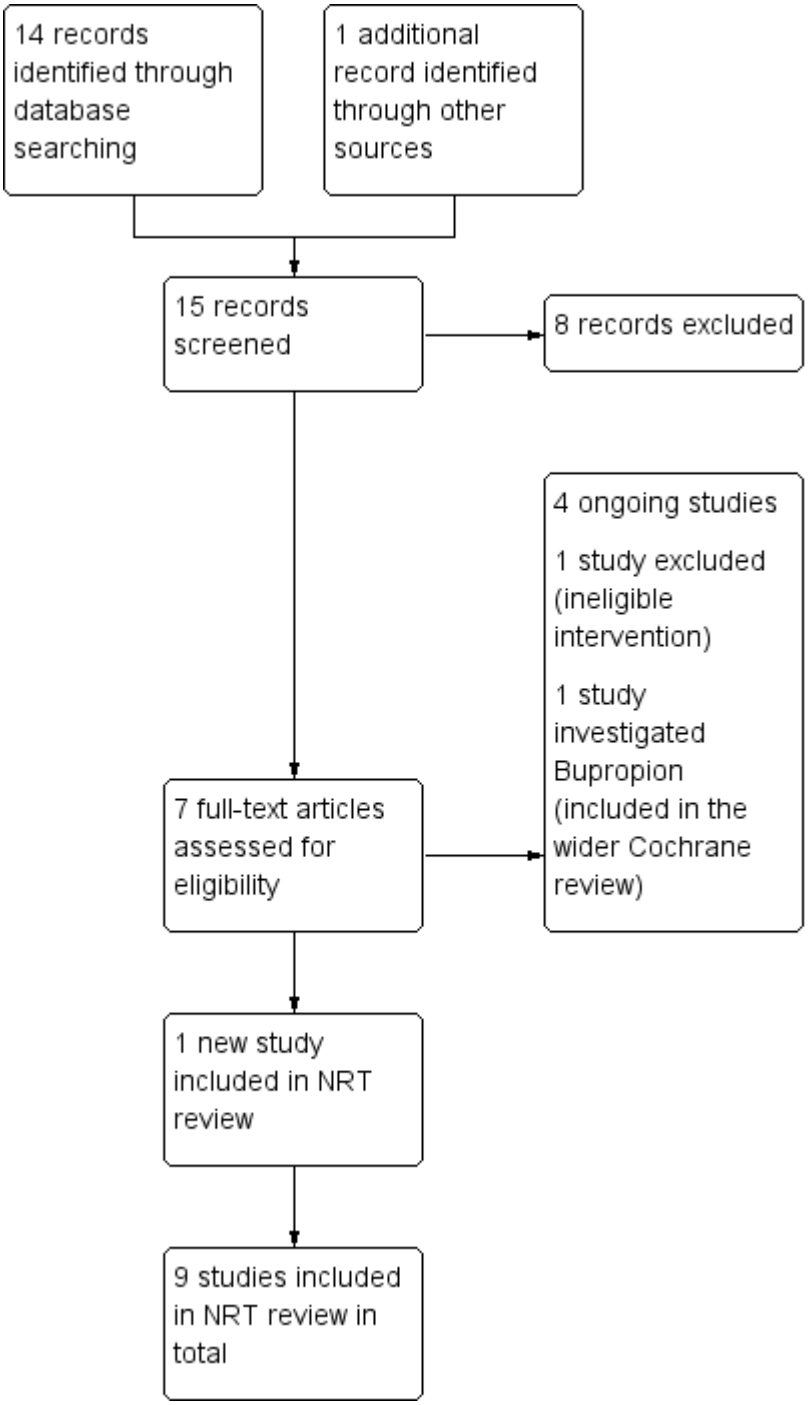

Figure 16 PRISMA flow diagram for updated review search.

5.4.2 Included studies

5.4.2.1 Interventions

Nine studies investigated the efficacy of different forms of NRT (51-58, 151).

5.4.2.2 Nicotine replacement therapy studies

All included studies investigated the efficacy of NRT provided with behavioural support and compared this with either behavioural support alone or support plus a placebo, therefore studies measured the effect of NRT provided as an adjunct to behavioural support. Six papers described placebo-controlled RCTs (51-55, 151). Three trials compared NRT plus behavioural support with behavioural support alone (56-58); thus, participants in these studies could not be blinded to treatment. Two studies used fast-acting NRT, one using nicotine gum (53), and the other nicotine inhalers (151); six trials used nicotine patches (51, 52, 54-57); and one offered a choice of NRT formulations: approximately two-thirds of participants chose patches, whilst the remainder elected to use gum and lozenges (58).

Oncken *et al*., 2008 (53) used 2 mg nicotine gum, and Oncken *et al*., 2019 (151) used 4 mg nicotine inhalers. Four studies used 15 mg/16-hour nicotine patches (52, 55, 57, 58); one of these used a higher nicotine dose (21 mg/24 hours removed at night) for participants who reported smoking more than 15 daily cigarettes (58). Two studies attempted to match nicotine doses prescribed with either saliva,(51), or urinary cotinine levels (56), obtained at earlier appointments. Depending on cotinine levels, women in one study were treated with combinations of 10 mg and 15 mg 16-hour patches (51), and in the other study with 21 mg, 14 mg, or 7 mg 24-hour patches, with instructions to remove these at night (56). One trial advised women to use trial treatments from randomisation until childbirth, irrespective of whether or not they had relapsed to smoking (51), and another trial encouraged continued use of treatment for six weeks as long as the woman was actively trying to quit smoking (151). Other trials advised

women to stop using NRT if they restarted smoking and had a defined period for use of NRT.

5.4.2.3 Setting

Studies were conducted in the USA (n = 4) (53, 56, 58, 151), Australia (n = 1) (57), Canada (n = 1) (54), Denmark (n = 1) (55), France (n = 1) (51), and England (n = 1) (52). All trials were conducted in public hospitals or antenatal clinics.

5.4.2.4 Outcomes

In one study, smoking cessation was ascertained between 20 and 28 weeks' gestation (54); however, in all other studies this was ascertained at 32 weeks or later. In all of the included studies, biological samples were obtained from participants, and after any required clarification from the authors we determined that all used such samples to validate reported cessation at the primary endpoint: four studies used exhaled carbon monoxide (53, 56, 57, 151); three saliva cotinine (51, 55, 58); and one used both exhaled carbon monoxide and saliva cotinine (52). One study reported both thiocyanate and cotinine concentrations (54). For two studies, cut points were obtained from the trial authors (55, 58), and we obtained further data on biochemical validation from the authors of a trial that used a higher-than-standard cut point for saliva cotinine (26 ng/mL) (55). This revealed that the cotinine assay used had a lower limit of 20 ng/mL, which was also above the currently accepted cut point of 10 ng/mL, so some women who smoke may have been wrongly categorised as abstinent in this study.

The periods of abstinence from smoking that participants were required to demonstrate varied across studies. For smoking outcomes measured at

delivery, three studies reported both seven-day point prevalence abstinence from smoking and a measure of continuous abstinence simultaneously (51, 52, 58); however, definitions varied. One study (52), permitted a small number of temporary lapses to smoking as recommended by the Russell Standard criteria for outcome measurement in smoking cessation studies (146). The remaining two studies did not permit temporary lapses and defined continuous abstinence as seven-day point prevalence abstinence recorded on three (58), or up to seven occasions (51). Four studies reported only seven-day point prevalence abstinence (53, 55, 58, 151), and three reported point prevalence abstinence for an unstated period (54, 56, 57). Four studies reported seven-day point prevalence abstinence data at time points after childbirth: Wisborg *et al*., 2000 (55) provided data at three and 12 months postnatally; Coleman *et al*., 2012 (52) at six, 12, and 24 months; Oncken *et al*., 2008 (53) at six to 12 weeks (biochemically validated data); and Pollak *et al*., 2007 (58) at three months. Additionally, Coleman *et al*., 2012 (52) reported continuous abstinence between a quit date and each time point, allowing for temporary lapses too. Two studies reported self-reported maternal smoking at 12 months after childbirth (52, 55).

Infant and fetal safety outcomes were reported in seven studies (51-53, 55, 56, 58, 151). All seven of these studies reported mean birthweight and mean gestation age at delivery, and all reported the incidences of low birthweight births (defined as below 2500 g). Six of these studies reported rates of preterm birth defined as born before 37 weeks' gestation (51-53, 55, 58, 151). Six studies reported rates of miscarriage/spontaneous abortion and stillbirth (51-53, 55, 58, 151), and four trials also reported infants' rates of special care admission and neonatal death (51-53, 58). Two trials reported data on maternal hypertension in pregnancy or measured arterial blood pressure at each visit (51, 52), three trials reported rates of

congenital malformation (51, 52, 151); and two of these three trials reported rates of caesarean section (51, 52). Two trials reported single and multiple pregnancy data together, but authors supplied data for singleton pregnancies separately (51, 58).

With regard to the pre-birth fetal outcomes of miscarriage/spontaneous abortion and stillbirth, Oncken *et al*., 2008 (53) reported that, within singleton pregnancies, three control group participants had terminations that were performed for social reasons (presumed healthy fetus), so these fetuses were removed from the denominator for control group analyses (control group n = 91). Also, Pollak *et al*., 2007 (58) reported one fetal death prior to randomisation that was documented by ultrasound scanning (i.e. a 'missed abortion') in the NRT group, so this fetus was removed from the denominator for the NRT group (NRT group n = 121). Coleman *et al*., 2012 (52) reported one termination and one fetal death prior to randomisation in women allocated to NRT, so these two cases were removed from the NRT group denominator (NRT group n = 515). Berlin *et al*., 2014 (51) reported one termination in each trial group, both of which were conducted for fetal abnormalities that were assessed as not being compatible with survival at birth. Consequently, as these terminations were undertaken at 25 (placebo group) and 32 weeks, they have been counted as stillbirths in the analysis and remained in the denominator as well.

Coleman *et al*., 2012 (52) additionally reported infants' "survival without developmental impairment" and respiratory symptoms at two years of age and self-reported maternal smoking at six and 24 months after childbirth.

5.4.2.5 Ongoing studies

One study reported as ongoing in the previous review has now completed, with results published, and is now an included study in this review (151).

Two further NRT studies were identified as ongoing (153, 154). One study appears to offer NRT as part of a multicomponent intervention (153), which would likely not be included in this review; however, we will wait for further information to become available before making a decision to exclude. The other study, based in Iran, is currently aiming to recruit 1050 pregnant women to a RCT testing 15mg/16-hour nicotine patches versus a placebo control (154). This study appears to be eligible for inclusion in any future updates of this review.

5.4.3 Excluded studies

We excluded one trial following full-text screening in this update (152). This was a pilot cluster-randomised step-wedge trial, where NRT was part of a multimodal intervention that provided educational resources to health providers at aboriginal medical services. We judged that due to the study design and the multimodal intervention strategy, it was not possible to identify the independent effect of NRT on smoking cessation from this study.

5.4.4 Risk of bias in included studies

We judged four of the nine included studies to be at low overall risk of bias (51-53, 55), three as at high risk of bias (56-58), and the remainder unclear risk of bias (**Figure 17**).

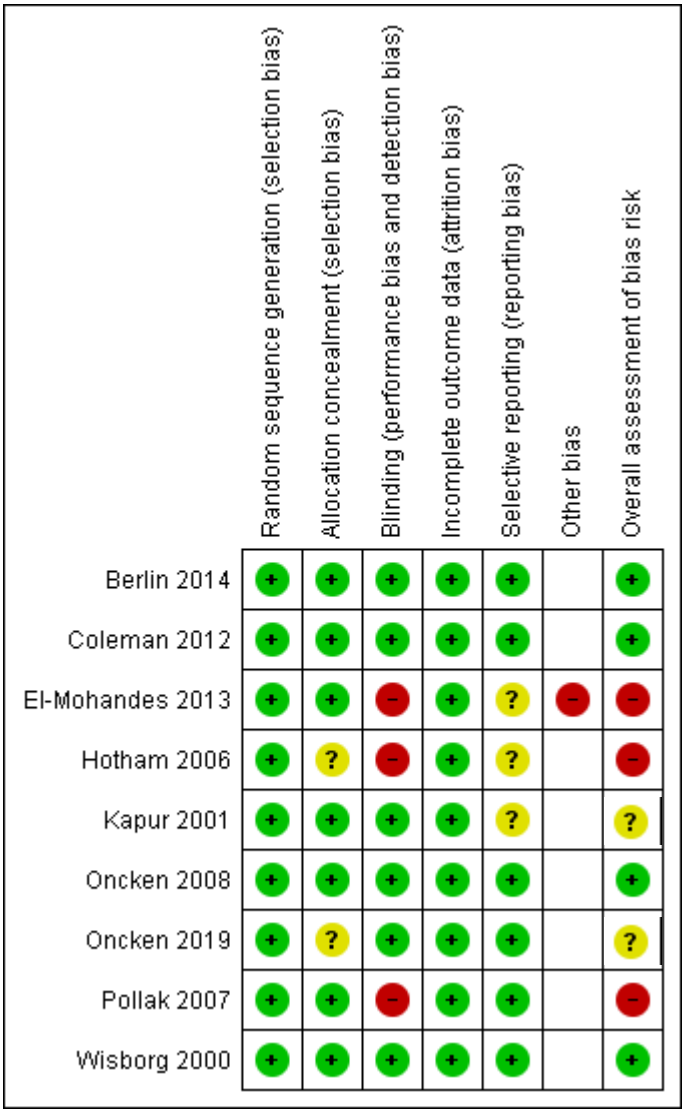

Figure 17 Methodological bias summary: review authors' judgements about each methodological bias item for each included study.

5.4.4.1 Allocation (selection bias)

Computer-generated random number sequences were used to generate randomisation in most studies. One study used urn randomisation (a method that is systematically based in favour of balancing of covariates, preserving randomization as the primary basis for assignment to treatment (155)) and was judged to be at low risk of bias for random sequence generation, but was unclear for allocation concealment due to insufficient detail (151). One study used sealed envelopes after random numbers had been generated, but it was not clear if these were opaque and sequentially

numbered (57); we therefore judged this study to be at unclear risk of bias for allocation, whilst the others were rated as satisfactory (low risk of bias).

5.4.4.2 Blinding (performance bias and detection bias)

We judged studies that had no placebo control to be at a high risk of bias, which was the principal difference between studies that was likely to cause bias. Six trials were placebo-controlled RCTs (51-55, 151), and three studies compared behavioural support alone with NRT and behavioural support (56-58).

As all included trials biochemically validated self-reported smoking outcomes, detection bias is not a major issue for this review. However, one included study used a cut point for saliva cotinine (26 ng/mL) that was substantially higher than the currently accepted level (10 ng/mL) and, additionally, used an assay with a lower limit of measurement of 20 ng/mL (i.e. samples in the 0 to 20 ng/mL range were reported as 20 ng/mL) (55). This means that some of the participants who may have falsely reported themselves as not smoking in this study might have had their false reports of abstinence validated as true (i.e. some participants who were actually smoking might not have had this detected by the validation process). Of course, no validation process is perfect, and, using any cut point, some false reports of cessation would be accepted to be true, but with a known high cut point as in Wisborg *et al*., 2000 (55), this would be expected to occur more frequently. However, the use of biochemical validation in this study would still be expected to detect heavier smoking in those who made false reports of abstinence, so validated data from this study were still used in preference to self-report data.

5.4.4.3 Incomplete outcome data (attrition bias)

We judged all studies to be at low risk of bias for smoking abstinence outcomes; all studies carried out an intention-to-treat analysis, so that those participants who could not be contacted at follow-up were assumed to have returned to smoking. It should be noted that this assumption is conservative and is the standard approach taken when assessing the efficacy of smoking cessation interventions. Follow-up for birth outcomes was generally high with one exception: the treatment group allocation for seven women who experienced miscarriage after being randomised within one study could not be ascertained (55); as this was not the primary outcome, we assessed this trial as at low risk of attrition bias.

5.4.4.4 Selective reporting (reporting bias)

We judged three studies as at unclear risk of reporting bias. Hotham *et al*., 2006 (57) collected data on a number of outcomes that were not reported in the trial manuscript; however, it is unclear whether this was a source of bias. We requested birthweight information from Hotham *et al*., 2006 (57) for our meta-analysis but were unable to obtain it. El-Mohandes *et al*., 2013 (56) informed us that within their trial, some data on secondary smoking cessation outcomes were collected, but this information was not reported in the trial manuscript; however, primary outcomes were reported. Kapur *et al*., 2001 (54) did not report any birth outcomes. We judged the remaining six studies to be at low risk of reporting bias.

5.4.4.5 Other potential sources of bias

We identified an unanticipated potential source of bias in one study (56): two participants were screened and randomised on two separate occasions, with each pregnancy counted as a discrete study participation, and both

women included in the trial analysis twice. We considered this as potentially introducing bias into what was a relatively small study, and so judged this study as at high risk of bias.

5.4.5 Effects of interventions

Data were not identified for all pre-specified outcomes. Where data were available this is summarised below.

5.4.5.1 Primary outcomes (efficacy)

In a pooled analysis of nine included studies and 2336 participants, we found evidence that the use of NRT, as an adjunct to behavioural support, may result in a clinically significant improvement in smoking cessation rates in later pregnancy relative to control (risk ratio (RR) 1.37, 95% confidence interval (CI 1.08 to 1.74; I^2 = 34%; **Figure 18)**.

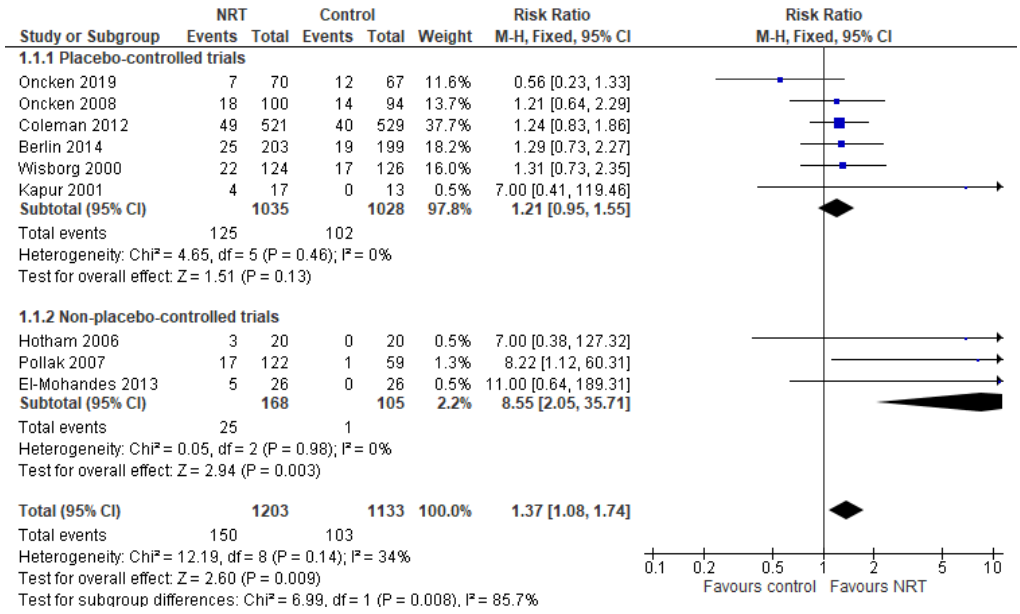

Figure 18 Forest plot of nicotine replacement therapy versus control, outcome: Validated cessation in later pregnancy (subgrouped by comparator type).

We carried out a subgroup analysis splitting the studies by comparator type - placebo or no placebo- and found evidence of a subgroup difference (P = 0.008; **Figure 18**). In the subgroup that compared active NRT with placebo, heterogeneity between studies was substantially reduced (I^2 = 0%), however the CIs incorporated the potential for both no effect and a benefit of NRT for smoking cessation (RR 1.21, 95% CI 0.95 to 1.55; 6 studies, 2063 women; **Figure 18**), whereas the estimate derived from non-placebo-controlled trials indicated only benefit (RR 8.55, 95% CI 2.05 to 35.71; I^2 = 0%, 3 studies; 273 women), but was limited by substantial imprecision. When analysing the data split into fast-acting and nicotine patch subgroups, the test for subgroup differences provided no evidence that the effect of NRT differed by type (P = 0.08; **Figure 19**).

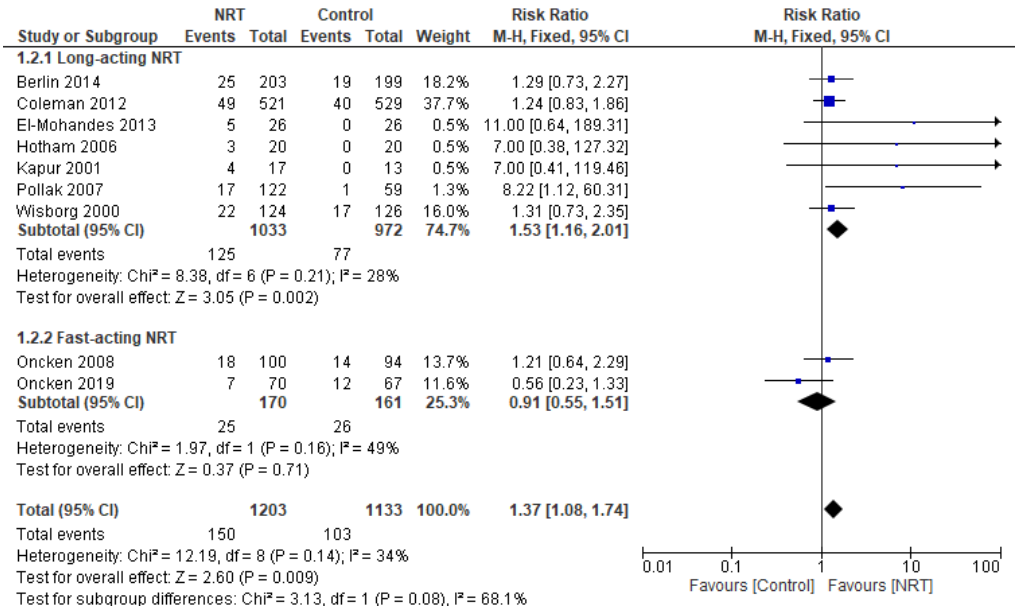

Figure 19 Forest plot of nicotine replacement therapy versus control, outcome: Validated cessation in later pregnancy (subgrouped by NRT type).

We planned to carry out a sensitivity analysis removing all studies judged to be at high risk of bias. The six studies that did not have a high risk of bias for any domain were the same six studies in the placebo-controlled

trials subgroup. This analysis and resulting 95% CI found evidence of potentially no clear effect of NRT, as well as the potential for benefit, therefore its interpretation does differ very slightly from that of the overall pooled analysis (**Figure 18**). We were unable to conduct the planned sensitivity analysis relating to adherence to treatment as trials reported adherence so differently that it was not possible to categorise one or more trials as having substantially worse or better treatment adherence than others.

We investigated the impact of NRT as an adjunct to behavioural support on cessation at time points after childbirth by pooling data from studies that provided postnatal follow-up data on smoking behaviour. In a pooled analysis of studies that reported non-validated seven-day point prevalence smoking abstinence up to six months after childbirth (predominantly at or around three months), there was no clear evidence that NRT compared to control was effective for smoking cessation, as CIs incorporated both potential benefit and harm of the intervention (RR 1.22, 95% CI 0.84 to 1.78; I^2 = 0%, 3 studies, 625 women; **Figure 20**). There was no statistical difference when comparing studies that were placebo controlled to the one study that was not (P = 0.59). Similarly, the pooled estimate for non-validated seven-day point prevalence smoking abstinence when comparing NRT to placebo at one year after childbirth resulted in CIs that incorporated both a small potentially negative effect of NRT, as well as a potentially positive effect at this time point (RR 1.35, 95% CI 0.97 to 1.88; I^2 = 5%, 2 studies, 1296 women; **Figure 21**).

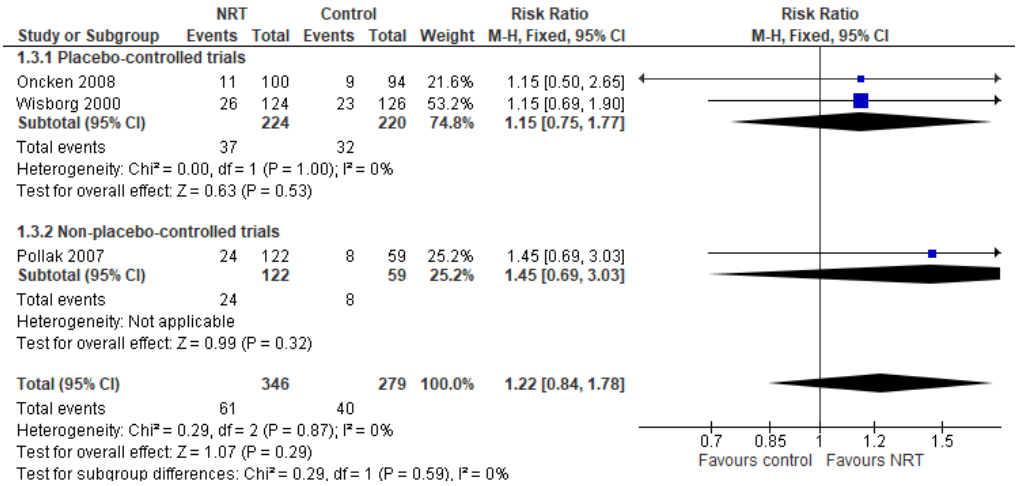

Figure 20 Forest plot of nicotine replacement therapy versus control, outcome: Self-report cessation at 3 or 6 months after childbirth

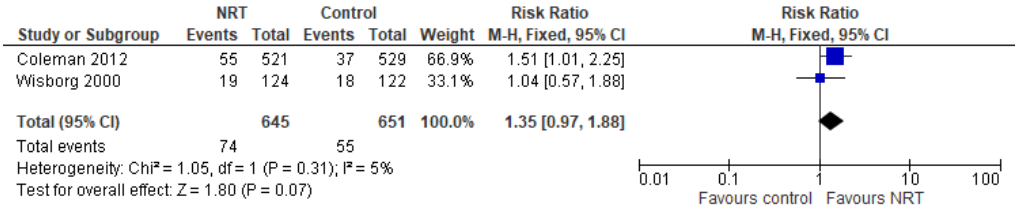

Figure 21 Forest plot of nicotine replacement therapy versus control, outcome: Self-report cessation at 12 months after childbirth

The one study that monitored continuous cessation from a quit date set in pregnancy to postnatal time points alongside seven-day point prevalence abstinence data collected at the same time points reported higher point prevalence than continuous cessation rates at each time point, and rates of continuous cessation until two years after childbirth were low (2.9% in the NRT group versus 1.7% in the placebo group, P = 0.20) (156).

5.4.5.2 Secondary safety outcomes

Two study papers reported birth outcomes from single- and multiple-birth infants together (51, 58); the authors kindly provided data on birth

outcomes within singleton pregnancies only to enable data from those studies to be included in the meta-analyses.

5.4.5.2.1 Miscarriage/spontaneous abortion

There was no evidence of a difference in risk of miscarriage/spontaneous abortion between the NRT and control group, and CIs incorporated the possibility of both potential benefit and harm of the intervention (RR 1.60, 95% CI 0.53 to 4.83; I^2 = 0%, 5 studies, 1916 women; **Figure 22**). However, despite contacting the study authors, we could not determine the treatment allocation for seven miscarriages from one study, which is not included in this comparison (55). If we assume that all miscarriages from this study occurred in either the NRT or the control group (i.e. the extremes of how these could actually be distributed), this results in the following effect estimates: all assumed in the NRT group: RR 2.15, 95% CI 0.77 to 6.02; all assumed in the control group: RR 1.06, 95% CI 0.38 to 2.97. This has no effect on the interpretation of the results.

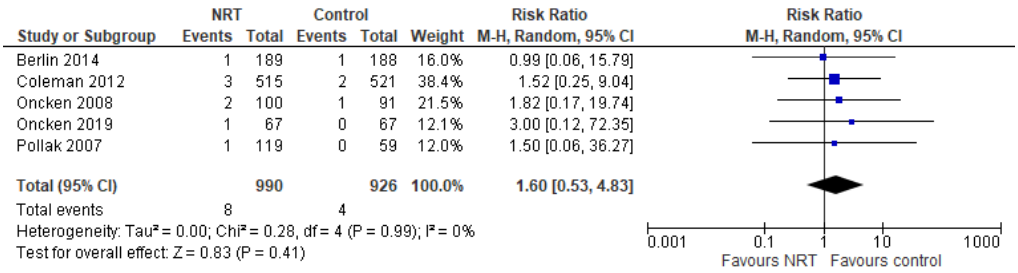

Figure 22 Forest plot of nicotine replacement therapy versus control, outcome: Miscarriage and spontaneous abortion

5.4.5.2.2 Stillbirth

Similarly, there was no evidence of a difference between the numbers of stillbirths in the NRT and control groups (RR 1.24, 95% CI 0.54 to 2.84; I^2 = 0%, 4 studies, 1777 women; **Figure 23**).

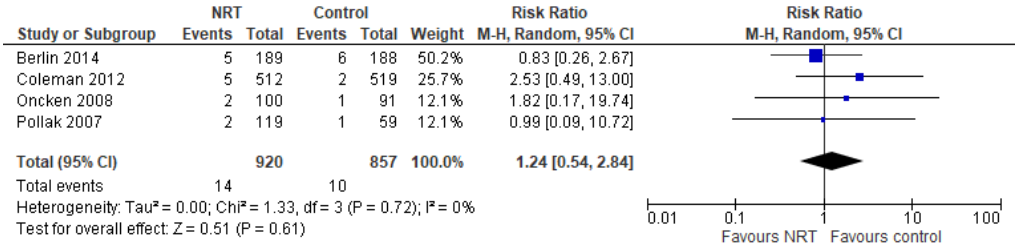

Figure 23 Forest plot of nicotine replacement therapy versus control, outcome: Stillbirth

5.4.5.2.3 Mean unadjusted birthweight

Despite the pooled estimate for birthweight being higher for the NRT group than for the control group, there was no evidence of a difference of mean birthweight between the NRT and control groups (mean difference (MD) 99.73 g, 95% CI −6.65 to 206.10; I^2 = 70%, 7 studies, 2202 women; **Figure 24**). Heterogeneity was high; the result for this comparison must therefore be interpreted with caution. The reasons for this heterogeneity are unclear; it is not easily explained by study design as one large placebo-controlled RCT (52), and a smaller non-placebo-controlled one (58), both reported non-significantly lower birthweight in NRT group infants, in contrast to other studies.

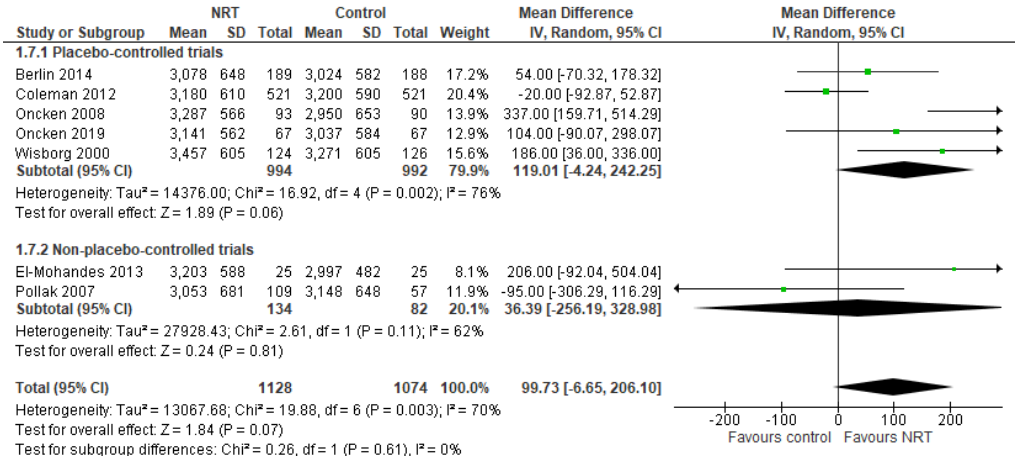

Figure 24 Forest plot of nicotine replacement therapy versus control, outcome: Mean birthweight (g)

5.4.5.2.4 Low birthweight (less than 2500g)

There was a lower incidence of low birthweight births in women in the NRT group, but again this was not significant and was found in the context of much heterogeneity, so caution is again warranted (RR 0.69, 95% CI 0.39 to 1.20; I^2 = 69%, 7 studies, 2171 women; **Figure S1**). The pattern of heterogeneity was once again difficult to understand: the same two studies reported non-significantly higher rates of low-birthweight infants in the NRT arm (52, 58).

5.4.5.2.5 Preterm birth (less than 37 weeks' gestation)

Analyses of rates of preterm births (RR 0.81, 95% CI 0.59 to 1.11; I^2 = 21%, 7 studies, 2182 women; **Figure S2**) resulted in CIs spanning one, incorporating the potential for both benefit and harm.

5.4.5.2.6 Neonatal intensive care unit admissions

There was no evidence of a difference in risk of neonatal intensive care unit admissions between the NRT and control groups (RR 0.90, 95% CI 0.64 to 1.27; I^2 = 0%, 4 studies, 1756 women; **Figure S3**).

5.4.5.2.7 Neonatal death

Similarly, there was no evidence of a difference in risk of neonatal deaths between the NRT and control groups (RR 0.66, 95% CI 0.17 to 2.62; I^2 = 0%, 4 studies, 1746 women; **Figure S4**).

5.4.5.2.8 Caesarean section

A meta-analysis of rates of caesarean birth suggested no clear evidence for a benefit or harm of NRT (RR 1.18, 95% CI 0.83 to 1.69, I^2 = 46%, 2 studies (51, 52), 1401 women; **Figure S5**).

5.4.5.2.9 Congenital anomaly

The same two studies that reported caesarean section, also reported congenital anomalies. The meta-analysis found no clear evidence for a benefit or harm of NRT (RR 0.73, 95% CI 0.36 to 1.48, I^2 = 0%, 2 studies (51, 52), 1401 women; **Figure S6**).

5.4.5.2.10 Maternal hypertension

The three studies that provided data on blood pressure (BP) reported these in different formats: Coleman *et al*., 2012 (52) reported that 24 (4.6%) in the NRT group compared to 25 (4.7%) in placebo were noted to have hypertension in pregnancy (i.e. BP of greater than 140/90 mmHg) on at least two occasions (no statistical comparison presented). Berlin *et al*., 2014 (51) reported significantly higher median diastolic BP in the NRT group

(median BP = 70, interquartile range (IQR) = 60 to 80 mmHg) compared to placebo (median BP = 62, IQR = 60 to 80 mmHg) (P = 0.02). Berlin *et al*., 2014 (51) also reported an interaction between treatment group and time (i.e. during pregnancy) for increases in diastolic BP, though absolute increases in BP were small.

5.4.5.2.11 Infant respiratory symptoms

Coleman *et al*., 2012 (52) and Berlin *et al*., 2014 (51) also reported the distribution of mechanical ventilation of infants between NRT and placebo groups; no statistically significant differences were noted.

5.4.5.2.12 Infant development

Coleman *et al*., 2012 (52) was the only included study that reported infant outcomes after the neonatal period. Using a composite, self-report outcome based on the Ages and Stages Questionnaire, 3rd edition (ASQ-3) instrument (157), significantly better infant developmental outcomes were observed in infants born to women who had been randomised to NRT compared to those in the placebo group. The odds ratio (OR) for infants reaching two years of age 'without developmental impairment' (i.e. normal development) was 1.40 (95% CI 1.05 to 1.86).

5.4.5.3 Adherence and adverse effects

Where adherence was reported, this was generally low, as the majority of participants in all studies did not use complete courses of the NRT offered (**Appendix A**). Berlin *et al*., 2014 (51) differed from other studies in that transdermal patches were offered to women at 3 time points between their quit dates and delivery, whereas other studies offered NRT once. Much higher self-reported adherence rates were noted in this study; however, it

is difficult to reconcile these with reported rates of intervention discontinuation, and direct comparison with other studies was not possible.

5.4.5.4 Non-serious adverse effects

Only a narrative reporting of non-serious adverse effect data was possible due to wide ranging effects. Six NRT trials reported non-serious adverse effects (51-53, 55, 57, 151). One trial reported their frequency within women using NRT, noting that five (25%) women in the NRT group experienced minor symptoms, and two women stopped using patches after unpleasant effects (57); however, non-serious adverse effects were not monitored in the control group, so this figure is difficult to interpret. Oncken *et al*., 2008 (53) reported that at least 10% of participants experienced headache, dizziness, fatigue, heartburn, nausea or vomiting, with 14 (15%) in the NRT and 12 (12%) in the control groups discontinuing treatment due to adverse effects. Wisborg *et al*., 2000 (55) noted that 11 women stated that adverse effects (e.g. skin irritations and headache) made them discontinue patches, but did not report treatment allocations; this trial also reported that five women experienced palpitations and two nausea. Coleman *et al*., 2012 (52) noted 535 non-serious adverse events reported by 521 NRT group participants and 450 reported by 529 placebo group participants. Berlin *et al*., 2014 (51) reported a range of non-serious adverse events, noting that more non-gynaecological ones occurred in the NRT group, but this was principally due to skin reactions. In this study, 11% of participants in the NRT group suffered a skin reaction at the patch site compared with 4% in the placebo group. Oncken *et al*., 2019 (151) reported a significantly higher number of adverse effects in women using the nicotine inhaler (11%) than the placebo inhaler (0%) ($P = 0.008$). These adverse events included throat irritation, cough, and nausea. Furthermore, two women in this study were discontinued from the nicotine inhaler group due

to repeated elevations in cotinine concentrations exceeding more than 40% of their baseline cotinine concentration.

5.5 Discussion

5.5.1 Summary of main results

Overall there is low-certainty evidence that NRT used alongside behavioural support by pregnant women for smoking cessation may increase smoking abstinence in late pregnancy (**Table 1**). Caution is required when interpreting this pooled estimate, as subgroup analyses revealed potentially different treatment effects when comparing NRT to placebo-controlled versus non-placebo-controlled studies. These findings may be due to unexplained biases potentially within the less robust, non-placebo-controlled trials. The actual efficacy of NRT used for smoking cessation in pregnancy is uncertain and may be lower than the pooled summary estimate (**Figure 18**). Further subgroup analysis found no evidence that the effect of NRT on abstinence is moderated by the type of NRT used, that is patches versus fast-acting NRT, and there was no consistent evidence of NRT having either a positive or negative impact on birth outcomes.

Table 1 Summary of findings table

Nicotine replacement therapy compared to control for smoking cessation during pregnancy

Patient or population: pregnant women who smoke
Setting: public hospitals and antenatal clinics (Australia, Canada, Denmark, France, the UK, the USA)
Intervention: nicotine replacement therapy
Comparison: placebo plus similar/matched behavioural support or similar/matched behavioural support only

	Anticipated absolute effects* (95% CI)					
Outcomes	**Risk with placebo/no NRT**	**Risk with NRT**	**Relative effect (95% CI)**	**№ of participants (studies)**	**Certainty of the evidence (GRADE)**	**Comments**
Biochemically validated smoking cessation at the latest point in pregnancy (20 weeks' gestation or more)	Study population 9 per 100	12 per 100 (10 to 16)	RR 1.37 (1.08 to 1.74)	2336 (9 RCTs)	⊕⊕⊖⊖ LOW [1] [2]	
Mean birthweight (g)	Study population 3139 g [3]	3239 g (3132 g to 3345 g)	MD 99.73 g (−6.65 g to 206.10 g)	2202 (7 RCTs)	⊕⊕⊖⊖ LOW [4] [5]	
Miscarriage and spontaneous abortion	Study population 0 per 100	1 per 100 (0 to 2)	RR 1.60 (0.53 to 4.83)	1916 (5 RCTs)	⊕⊕⊖⊖ LOW [6]	

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: confidence interval; **MD:** mean difference; **NRT:** nicotine replacement therapy; **RCT:** randomised controlled trial; **RR:** risk ratio

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect.
Moderate certainty: We are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low certainty: Our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
Very low certainty: We have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

Footnotes

[1]Downgraded one level due to serious risk of bias: in the subgroup of studies at low or unclear risk of bias the effect was no longer statistically significant, and there were significant subgroup differences when comparing these studies to the three studies judged to be at high risk of bias (P = 0.008).
[2]Downgraded one level due to serious imprecision: there were only 253 events in total (300 to 400 recommended for dichotomous outcomes), and confidence intervals span both minimal clinical benefit and considerable clinical benefit.
[3]Control risk based on observed birthweights in the control arms.
[4]Downgraded one level due to serious inconsistency: I^2 = 70%, not explained by subgroup differences.
[5]Downgraded one level due to serious imprecision: confidence intervals encompass no difference as well as a clinically significant benefit.
[6]Downgraded two levels due to very serious imprecision: there were only 12 events in total (300 to 400 recommended for dichotomous outcomes), and confidence intervals encompass both no difference and potential harm.

5.5.2 Overall completeness and applicability of evidence

All of the included studies were conducted in high-income countries, with only one study specifically recruiting women from ethnic minority backgrounds. These findings may therefore not be applicable to low-middle-income countries if smoking patterns of women or beliefs about using medication in pregnancy differ, and more evidence is needed from these populations.

An exclusion criterion for this review was unmatched additional intervention components in the intervention or comparator arms. This means that we can be confident that we have isolated the independent effects of the interventions of interest to our review question.

It has been mandatory since July 2005 for clinical trials to be recorded on a trials register. In this update we searched trials registers from inception, therefore we are confident that we have identified all reported ongoing trials.

The findings reported in this review are based on currently accepted, evidence-based, biochemical verification cut points for determining abstinence from smoking (158), rather than ones that might have been acceptable in the past, enhancing the validity of our findings.

5.5.3 Certainty of the evidence

The included trials had varied 'Risk of bias' ratings (**Figure 17**). We assessed four of the nine included studies to be at low risk of bias, three at high risk of bias, and the remainder at unclear risk of bias. We judged the principal difference in studies' propensity to bias to be due to the use/non-use of placebo controls. The reduction in heterogeneity observed after grouping trials according to this criterion seemed to validate this judgement.

Trials that were judged to be at an unclear risk of bias lacked information regarding allocation concealment or did not report prespecified outcomes. It is possible, but relatively unlikely, that the lack of information regarding allocation concealment indicates bias.

We assessed the certainty of the evidence using the GRADE approach for critical and important outcome measures. The GRADE assessment of pooled data indicated that the evidence for the smoking cessation outcome in NRT trials was of low certainty (**Table 1**), meaning that the true effect might be markedly different from the estimated effect. The current evidence was downgraded twice, once due to risk of bias: in the subgroup of studies at low or unclear risk of bias the effect was no longer statistically significant, and there were significant subgroup differences when comparing these studies to the three studies judged to be at high risk of bias. We downgraded the evidence further due to serious imprecision, as there were few events, and confidence intervals spanned both minimal clinical benefit and considerable clinical benefit. Both of these downgrades are subjective and could be considered marginal, however after discussion with other reviewers it was decided that these downgrades were justified. We assessed the evidence for the safety outcomes in NRT trials, mean birthweight and miscarriage, to be of low certainty. The mean birthweight outcome was downgraded due to inconsistency where heterogeneity was high and not explained by subgroup differences, and was further downgraded due to imprecision, as the pooled confidence interval encompassed no difference as well as a clinically significant benefit. Additionally, standard deviations were relatively large for most studies. The miscarriage and spontaneous abortion outcome was downgraded two levels to low certainty due to imprecision, as there were too few events, and confidence intervals encompassed both no difference and potential harm.

The downgrading of the evidence for all outcomes due to imprecision suggests that further research will be beneficial in increasing the reliability and precision of effect estimates and the certainty we are able to place in them.

5.5.4 Potential biases in the review process

We performed the search for studies in this area using the Cochrane Pregnancy and Childbirth Group's Trials Register. It is unlikely that studies that have been conducted have been missed, however it is possible that unpublished studies, or ongoing studies not registered in clinical trial registries, could be missing. Should we identify any such studies, we will include them in future updates of the review. Secondly, we were unable to produce a funnel plot as there were too few studies, and it is possible there was publication bias. In future updates where there are sufficient trials we will be able to assess publication bias more rigorously. Finally, we aimed to reduce bias wherever possible by having at least two review authors independently conduct study selection, data extraction, and 'Risk of bias' assessment.

5.5.5 Agreements and disagreements with other studies or reviews

This review explicitly assesses the efficacy and safety of pharmacological therapies used for smoking cessation in pregnancy. Some trials of smoking cessation in pregnancy test NRT as part of multimodal intervention strategies, and these are included in an associated review (40). However, this review was concerned with the efficacy and safety of NRT when used for smoking cessation in pregnancy, and examines the independent safety and efficacy of NRT.

We have been unable to identify any other systematic reviews that investigate the efficacy of smoking cessation medications in pregnancy since the previous version of this review was published (144). A systematic review of trials conducted in non-pregnant women has shown that NRT is effective outside of pregnancy (159). The reasons why NRT may not be as effective in pregnancy are not known; however, variations in adherence to NRT or nicotine metabolism compared to the general population may play a part. Women in trials included in the current review made relatively little use of offered NRT. If this low adherence explains the difference in findings between this and the 'non-pregnancy' NRT review (159), then understanding the phenomenon of low adherence could be important. Lack of efficacy could also be explained by the increased metabolism of nicotine in pregnancy (160). This may result in NRT generating lower blood nicotine concentration in pregnancy, and this reduced nicotine substitution could, in turn, increase women's experience of withdrawal symptoms, causing them to stop NRT early. A recent systematic review found that pregnant women using NRT were exposed to significantly lower concentrations of nicotine compared to those who continued to smoke tobacco (161). Furthermore, a secondary analysis of a trial included in our review found that pregnant women who both smoke and use nicotine patches had similar cotinine concentrations, smoke less, and exhale less carbon monoxide, therefore they are likely to be exposed to fewer tobacco smoke toxins (162). An increased metabolism of nicotine during pregnancy results in lower exposure, and coupled with the likelihood that nicotine is unlikely to be responsible for the majority of fetal harms caused by tobacco smoke, it is likely that NRT is safer for the fetus than smoking (163). Logically, if in trials to date, increased metabolism underpinned women's low adherence to NRT, higher doses of NRT could be needed for this to be effective in pregnancy.

5.6 Authors' conclusions

5.6.1 Implications for practice

The evidence suggests that nicotine replacement therapy (NRT) may be effective for smoking cessation in pregnancy, however there is uncertainty surrounding this evidence. It is also unclear whether NRT affects the risk of adverse pregnancy and infant outcomes, but there is no evidence that it is harmful. One study suggests that NRT improves child development outcomes at two years.

5.6.2 Implications for research

As adherence to NRT in pregnant women is low, further research should seek to understand why this is and improve it and use an appropriate behavioural strategy to enhance adherence in future trials of NRT. Qualitative studies could add further context as to why adherence to NRT in pregnancy is low.

In the general population, there is evidence that 25 mg/16-hour patches are more effective than 15 mg/16-hour patches (164); most studies in this review used 15 mg patches. Consequently, trials are needed in pregnant women using either higher-dose nicotine patches or combination of patch plus rapid-acting forms of NRT, which are also more effective (164).

There is a strong case for further trials to examine the effectiveness and safety of NRT against placebo. NRT leads to lower blood nicotine concentrations than when smoking and is effective in the general population (159), however there are also reasons why it may be less effective for pregnant women than for the general population, and the evidence in pregnant women is uncertain. The following chapter discusses whether conducting more trials is futile, and if not, how many more participants

would be needed in further trials to ascertain whether NRT is an effective treatment for smoking cessation in pregnancy.

Chapter 6: Trial Sequential Analysis of the efficacy of nicotine replacement therapy for smoking cessation during pregnancy

6.1 Introduction

In **Chapter 5** a systematic review and meta-analysis was performed to assess the efficacy of NRT used during pregnancy for smoking cessation. This review found that NRT use, together with behavioural support, is 37% more effective for smoking cessation in pregnancy relative to control (RR 1.37, 95% CI 1.08 to 1.74). However, subgroup analysis of only placebo-controlled studies found a more conservative estimate of effect, which is deemed not significant by traditional measures of statistical significance (RR 1.21, 95% CI 0.95 to 1.55). From these traditional meta-analyses, it is unknown whether the intervention effects are spuriously overestimated (type I error) or spuriously underestimated (type II error) due to insufficient randomised participants (165). **Chapter 3** introduces a method called trial sequential analysis which may overcome these issues. In this chapter we apply trial sequential analysis methods to the meta-analyses conducted in **Chapter 5**, for the primary outcomes in the Summary of findings table (biochemically validated smoking cessation at the latest point in pregnancy; mean birthweight; and miscarriage and spontaneous abortion).

6.2 Aim

The aim of this chapter is to determine whether there is sufficient information in meta-analyses regarding the efficacy and safety of NRT for smoking cessation in later pregnancy in **Chapter 5**.

6.3 Objectives

The study aim was investigated through the following objectives:

I. To use trial sequential analysis to assess whether NRT is effective for smoking cessation in pregnancy.

II. To use trial sequential analysis to assess whether NRT is safe for smoking cessation in pregnancy.

III. If unclear, to use TSA to discover whether conducting future trials is futile, or how many participants would be required in future studies to arrive at a firm conclusion.

6.4 Methods

TSA of a meta-analysis of RCTs in an analogous approach to interim analysis of a single RCT (166). TSA increases the uncertainty of an intervention effect if the cumulative information in the meta-analysis unsuccessfully achieves the minimum number of randomised participants to detect or reject a pre-specified clinically important effect size (104). This uncertainty is reduced if the proportion of randomised participants is higher in relation to the required information size. When the required information size has not been reached, the results from the TSA are adjusted to reflect the uncertainty through using TSA-adjusted CIs. Thus, the further the number of randomised participants included in the meta-analysis are from the required information size, the wider the TSA-adjusted confidence intervals. This means that the significance level is lower to assess the uncertainty of the point estimate (104). The required information size is calculated using the anticipated event proportion in the control group, a pre-specified plausible relative risk reduction or increase in the intervention group, and the anticipated heterogeneity variance (D^2) of the meta-analysis.

In this study, TSA was applied to the primary outcome of biochemically validated smoking cessation at the latest point in pregnancy, and the subgroup analysis by comparator. Trials were included sequentially based on date of publication according to the year of publication, and if more than one trial had been published in a year, we added these trials alphabetically

according to the last name of the first author. The required information size was estimated based on the control event proportion from the meta-analysis; D^2 as suggested by the meta-analysis; an alpha (type I error) of 5%; and beta (type II error) of 90%; and an anticipated relative risk reduction of that observed in trials with a low risk of bias. These parameters were decided *a priori,* as reported in the PROSPERO record (**Appendix B)**.

Sensitivity TSA were conducted using the same parameters as above, but instead using a D^2 of the upper bound of the 95% confidence interval for heterogeneity calculated by the TSA software.

6.5 Results

6.5.1 Trial sequential analysis for efficacy

Nine studies, including 2336 participants, reported data on smoking cessation at the latest time-point in pregnancy. The meta-analysis found evidence that the use of NRT, as an adjunct to behavioural support, may result in a clinically significant improvement in smoking cessation rates in later pregnancy relative to control (RR 1.37, 95% CI 1.08 to 1.74; I^2 = 34%; **Figure 18**). However, this result was not confirmed by TSA (TSA adjusted CI 0.52 to 3.64; **Figure 25, Table 2**). TSA analysis found that although the Z-curve crossed the conventional significance boundary (P=0.05) indicating a significant result for NRT, the curve did not cross TSA monitoring boundaries, demonstrating potentially early spurious results. The TSA reported that a further 18,708 participants from at least one additional trial would be needed to reach a firm conclusion regarding the effectiveness of NRT for smoking cessation in late pregnancy.

6.5.1.1 Sensitivity analysis

Sensitivity analysis, performing TSA using the upper bound of the 95% CI for D^2 (54%), was consistent with the primary analysis (TSA adjusted CI 0.52 to 3.64; **Table 2**). The required information size increased to 22,860, meaning that a further 20,524 participants from at least one additional trial would be needed to reach a firm conclusion regarding the effectiveness of NRT for smoking cessation in late pregnancy.

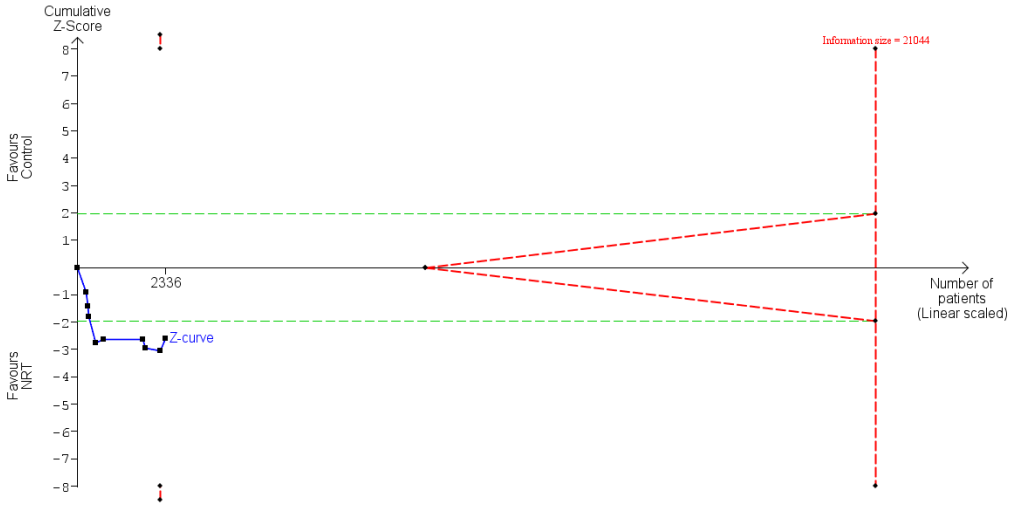

Figure 25 Trial sequential analysis of biochemically validated smoking cessation at the latest time point in pregnancy compared to placebo or non-placebo control. The required information size was calculated using α = 0.05, β = 0.90, relative risk reduction = based on low biased trials, diversity (D^2) as suggested by trials, and a control event rate of 9.09%. The cumulative Z-curve was constructed using a fixed-effects model, and each cumulative Z-value was calculated after inclusion of a new trial (represented by black dots). The horizontal green lines represent the conventional naïve boundaries for benefit. The etched lines represent the trial sequential boundaries for benefit, harm, or futility (middle triangular area). The cumulative Z-curve does not cross the TSA boundary for benefit, indicating future trials are required. The estimated information size is 21,044, meaning future trials would need approximately 18,708 participants in total for a firm conclusion.

6.5.2 Placebo-controlled trials subgroup analysis

Six studies with 2063 participants reported data on smoking cessation in pregnancy with a placebo comparator. These studies were deemed at lower risk of bias. However, in the conventional meta-analysis, the CIs

incorporated the potential for both no effect and a benefit of NRT for smoking cessation (RR 1.21, 95% CI 0.95 to 1.55; **Figure 18**). This result was confirmed by TSA (TSA adjusted CI 0.66 to 2.22; **Figure 26**, **Table 2**). In this scenario, TSA analysis found that the Z-curve did not cross either the conventional significance boundary or the TSA monitoring boundaries. However, futility boundaries were not crossed, meaning performing further trials would not be futile. Therefore, a further 8,453 participants from at least one additional placebo-controlled trial would be needed before a firm conclusion regarding the effectiveness of NRT for smoking cessation in pregnancy can be determined.

6.5.2.1 Sensitivity analysis

Sensitivity analysis, performing trial sequential analysis using the upper bound of the 95% CI for D^2 (49%), was consistent with the primary analysis (TSA adjusted CI 0.44 to 3.30; **Table 2**). The required information size increased to 20,619, meaning that a further 18,556 participants from at least one additional trial would be needed to reach a firm conclusion regarding the effectiveness of NRT for smoking cessation in late pregnancy.

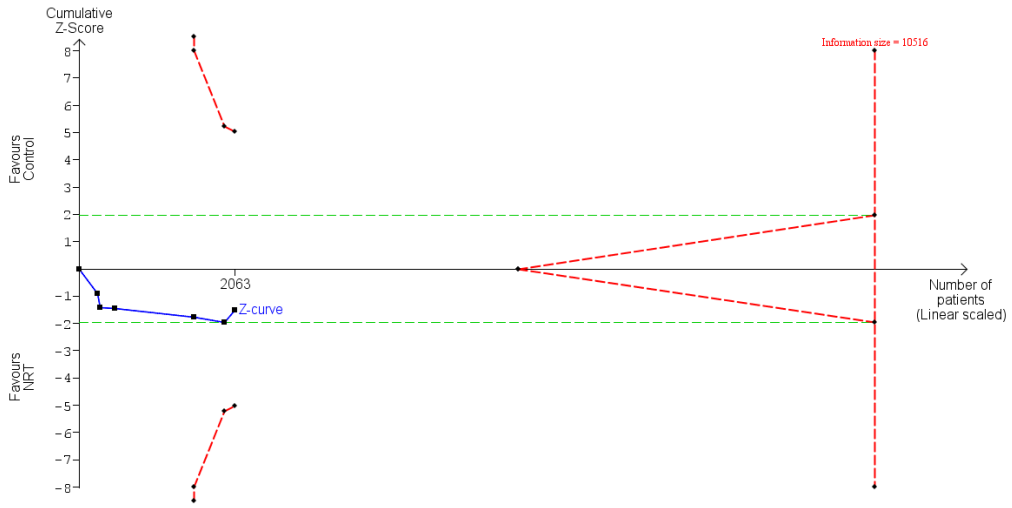

Figure 26 Trial sequential analysis of biochemically validated smoking cessation at the latest time point in pregnancy compared to placebo control

only. The required information size was calculated using α = 0.05, β = 0.90, relative risk reduction = based on low biased trials, diversity (D2) as suggested by trials, and a control event rate of 9.09%. The cumulative Z-curve was constructed using a fixed-effects model, and each cumulative Z-value was calculated after inclusion of a new trial (represented by black dots). The horizontal green lines represent the conventional naïve boundaries for benefit. The etched lines represent the trial sequential boundaries for benefit, harm, or futility (middle triangular area). The cumulative Z-curve does not cross the TSA boundary for benefit, indicating future trials are required. The estimated information size is 10,516, meaning future trials would need approximately 8,453 participants in total for a firm conclusion.

6.5.3 Trial sequential analysis for safety

6.5.3.1 Miscarriage and spontaneous abortion

Five studies with a total of 1916 participants reported miscarriage and spontaneous abortion as an outcome measure. Traditional meta-analysis found no evidence of a difference in risk of miscarriage/spontaneous abortion between the NRT and control group, and CIs incorporated the possibility of both potential benefit and harm of the intervention (RR 1.60, 95% CI 0.53 to 4.83; **Figure 22**). This result was confirmed by TSA analysis, however the adjusted confidence intervals were very wide (TSA adjusted CI 0.02 to 145.85; **Table 2**). In this TSA analysis the z-curve again did not cross either the conventional significance boundary or the TSA monitoring boundaries. However, futility boundaries were not crossed, meaning performing further trials would not be futile. In this circumstance sensitivity analysis was not possible, as using the upper bound of the 95% CI for D^2 (54%) meant that less than 5% of the required information size was accrued.

6.5.3.2 Mean birthweight

Seven studies incorporating 2202 participants reported mean birthweight as an outcome measure. In traditional meta-analysis, the pooled estimate for birthweight was higher for the NRT group than for the control group, but the CIs incorporated a small decrease in birthweight as well as a more

substantial increase (MD 99.73g, 95% CI −6.65 to 206.10; **Figure 24**). Trial sequential analysis widen the CIs further (TSA adjusted CI -64.65 to 264.11; **Table 2**). Like the output for miscarriage and spontaneous abortion, futility boundaries were not crossed, so further trials would not be futile in this instance. Sensitivity analysis using the upper bound of the 95% CI for D^2 (80%), further widen the confidence intervals (TSA adjusted CI -74.73 to 274.18; **Table 2**).

Table 2 Conventional meta-analysis and trial sequential analysis outcomes.

			Primary TSA[1]		Sensitivity TSA[1]	
Outcome	Number of trials (Participants)	Conventional meta-analysis	TSA adjusted 95% CI	Information size	TSA adjusted 95% CI	Information size
Smoking cessation	9 (2336)	RR 1.37 (95% CI 1.08 to 1.74)	0.52 to 3.64	21004	0.52 to 3.64	22860
Smoking cessation (placebo-controlled trials only)	6 (2063)	RR 1.21 (95% CI 0.95 to 1.55)	0.66 to 2.22	10516	0.44 to 3.30	20619
Miscarriage and spontaneous abortion	5 (1916)	RR 1.60 (95% CI 0.53 to 4.83)	0.02 to 145.85	34623	Insufficient data (<5% of IS)	Insufficient data (<5% of IS)
Mean birthweight	7 (2202)	MD 99.73g (95% CI −6.65 to 206.10)	-64.65 to 264.11	9669	-74.73 to 274.18	10970

[1]For dichotomous outcomes: α 5%; ß 90%; RRR low risk of bias based; D^2 model variance based. For continuous outcomes: α 5%; ß 90%; RRR low risk of bias based; D^2 upper confidence interval based.
α: two-sided significance level, ß: power; D^2: diversity; CI: confidence interval; IS: information size; RR: relative risk; MD: mean difference RRR: relative risk reduction; TSA: trial sequential analysis.

6.6 Discussion

According to the findings from the TSA, the current evidence from nine trials on the use of NRT during pregnancy is not sufficient to assess whether it aids smoking cessation during pregnancy compared to control. To reach a firm conclusion, a further 8,453 participants from at least one additional placebo-controlled trial is required.

As discussed in **Section 3.5.5**, the Cochrane Collaboration evaluated and updated their guidance on using TSA approaches in meta-analysis in their reviews (136, 137). The authors from the Cochrane Handbook for Systematic Review of Interventions concluded that TSA methods should not be used in primary analyses or to draw conclusions, but could be used as secondary analyses in reviews if they are prospectively planned and the assumptions underlying the design are clearly justified (137). In this case, all TSA analyses were prospectively planned, and parameters for the TSA were decided *a priori*. Additionally, the results from the TSA have been written up to avoid drawing binary conclusions and have not been influenced by plans for future updates (137).

To overcome methodological limitations of TSA methods when heterogeneity is present, a sensitivity analysis was performed using D^2 of the upper bound of the 95% confidence interval for heterogeneity calculated by the TSA. The findings of this sensitivity analysis show that the numbers increase substantially when substantial heterogeneity is present and therefore the estimate of heterogeneity (tau) needs to be robust. As there were only few studies included in the meta-analyses the estimate of tau is imprecise, which lead to the large increase in required information size. Although very impractical in the scenario above, it is important to consider whether to conduct one large multicentre trial or a number of smaller trials

to reach information size. It may be more insightful to conduct a series of smaller trials, as with more studies included in the meta-analysis, the estimate of tau will be more precise.

6.6.1 Conclusions

The results from the TSA suggest that further placebo-controlled trials comprising of a total of around 8,500 participants may be required to arrive at a stronger conclusion surrounding NRT use for smoking cessation in pregnancy. However, this figure may be impractical. In a period of over 18 years, only 2,083 women have been recruited to placebo-controlled trials. Thus, substantial time and resources would be necessary to recruit four times that amount. Furthermore, funders are unlikely to want to pay for such large studies to be conducted, especially if the traditional meta-analysis suggests that the intervention is likely to be effective.

Instead of focussing on NRT and how it is trialled currently, resources may be better spent understanding why NRT does not work as effectively during pregnancy. **Chapter 7** explores one such reason – by investigating concomitant smoking and NRT use and how this affects indicators of smoking intensity in pregnant women.

Chapter 7: Saliva cotinine concentrations in pregnant women who smoke and use nicotine patches

7.1 Introduction

The arguments presented in this chapter have been published in *Addiction* (**Appendix B**).

Smoking during pregnancy is the leading modifiable risk factor for poor maternal and infant health outcomes. Pregnancy-related health problems associated with smoking during pregnancy include complications during labour, increased risk of miscarriage, premature birth, stillbirth and low birth-weight (9, 11, 15). Despite this, around 12% of pregnant women in the UK, 13% in the United States and 20% in France continue to smoke during pregnancy (128, 167, 168). Several national guidelines have adopted using nicotine replacement therapy (NRT) for supporting pregnant smokers to quit, based on the idea that NRT is probably safer than smoking as it does not contain the toxins present in tobacco smoke (138, 169).

Whilst NRT has been proven to be effective in non-pregnant smokers (49), its efficacy in pregnancy is uncertain (59). It is unclear why the evidence for efficacy is uncertain, however it is hypothesised that physiological changes in pregnancy could affect nicotine's metabolism (170). Potential factors for the increased metabolism rate include a higher level or activity of metabolic enzymes involved and increased blood flow through the liver during pregnancy (68). Cotinine is the principal metabolite of nicotine, and the clearance of nicotine and cotinine is 60% and 140% higher respectively, during pregnancy (69). An increase in metabolic rate could signify that nicotine supplied through standard dose NRT may be insufficient to alleviate smoking withdrawal symptoms in pregnancy and to provide therapeutic effects.

A systematic review and meta-analysis comparing nicotine exposure in pregnant women when smoking, and their nicotine exposure when abstinent

and using NRT, found that NRT exposes women to lower doses of nicotine than smoking does (161). Generally, in studies included in this review, such as the Smoking, Nicotine and Pregnancy (SNAP) trial, women were instructed to discontinue use of nicotine patches if they had even brief smoking lapses (52). This mimics routine health care, where pregnant women are usually advised to stop using NRT if they lapse to smoking, even for short periods. There is concern that concomitant smoking and NRT use could increase exposure to nicotine and potentially more tobacco smoke toxins if they smoked heavily when using NRT. However, in pregnancy this assumption is untested, and we know little about women's smoking behaviour when they use NRT concurrently. This is important as women who lapse to smoking, may still want to quit. In a non-pregnant population continued use of nicotine patches has been found to promote recovery from lapses (171), if this is the case during pregnancy, women may have better chances of cessation if NRT is continued.

7.2 Aim

The aim of this chapter is to investigate the differences in indicators of smoking intensity in pregnant women when smoking before using NRT, and when using NRT and smoking concurrently.

7.3 Objectives

This study aim was investigated through the following objectives:

I. To investigate 'within-participant' differences in indicators of smoking intensity at two different time points in pregnancy, in women using patches and smoking concurrently, compared with those when they only smoked

II. To investigate if these changes differed between nicotine and placebo patch use.

III. To investigate whether an interaction between indicators of smoking intensity and nicotine patch assignment exist.

7.4 Methods

7.4.1 Design

This is a secondary analysis of data from the 'Study of Nicotine Patch in Pregnancy' (SNIPP) (51). SNIPP was a multi-centre, double-blind, randomised, placebo-controlled study conducted in France using 16-hour nicotine patches. The trial randomised 402 women to either nicotine (n=203) or placebo patches (n=199). The study was approved by the Ethics Committee of the Pitié-Salpêtrière Hospital, Paris, France.

7.4.2 Participants

Participants were eligible for inclusion in the SNIPP trial if they smoked at least 5 cigarettes per day, were aged over 18 years, of 12-20 weeks gestation and scored at least 5 on a scale measuring motivation to stop smoking (range 0-10) (51). Prior to enrolment, participants attended a baseline visit, where demographic, obstetric, physiological characteristics and smoking behaviour data were collected, and saliva cotinine concentrations were determined. At this stage, participants were given two weeks to quit smoking or reduce the number of cigarettes to fewer than five a day. If after this two-week period they were unable to do either of these, they could be randomised, receive the study drug, and set a quit date when treatment began. Participants were asked to stop smoking on a predefined quit date and were randomised to either placebo or nicotine patches. Participants were told that they could continue using nicotine patches during smoking lapses. Moreover, patch doses were adjusted according to the pre-

quit saliva cotinine determination to optimize the nicotine substitution; this resulted in participants receiving a mean nicotine dose of 18 mg/day (SD=6.8) in the nicotine patch arm.

7.4.3 Measures

In the SNIPP trial, abstinence was defined as self-reported abstinence, confirmed by expired air carbon monoxide concentration ≤8 parts per million (ppm) (Smokeanalyzer®, Bedfont Scientific Ltd, Rochester, Kent, UK) (172). Saliva cotinine samples were collected by placing a cotton roll in the gingival cleft for 1 minute, which was then placed immediately into a Salivette tube (Sarstedt, Nümbrecht, Germany) (172). Samples were kept at 4 °C and were sent to the central biochemistry laboratory (Hôpital Pitié-Salpêtrière, Laboratoire de Biochimie, Dr. N. Jacob) within 24 hours for determination (172). The quantification limit for cotinine was 7.5 µg/L and the between-run coefficient of variation 5–8% (172).

Figure 27 shows when trial visits occurred and when measurements were made. Saliva cotinine concentrations were determined at baseline, 2 weeks after quit date and 8 weeks after quit date, with nicotine doses adjusted after each of these visits at 4 weeks and 12 weeks after quit date respectively. Nicotine doses were adjusted using a conversion factor of 0.1. For example, a saliva cotinine concentration of 100 µg/L equated to a prescription of one 10mg patch (51). At baseline, body mass index (BMI), gestational age, ethnicity and Fagerström Test for Cigarette Dependence (FTCD) scores were recorded. As well as at baseline; at each visit, women reported any smoking in the previous week validated by expired air carbon monoxide. Additionally, intensity of craving for tobacco via the French Tobacco Craving Questionnaire, 12 items (FTCQ-12) and the number of cigarettes smoked by the participant in the last week were assessed. The

SNIPP trial recorded cigarette consumption in the past week, rather than cigarettes per day, due to large day to day fluctuations in cigarette consumption (173, 174). Partner smoking in the previous week was also assessed, as the second hand smoke exposure is likely to increase cotinine measures. Women were permitted to use nicotine patches from quit date up until delivery. A more extensive description is available elsewhere (51).

In this study we used data from women collected at 2-weeks after the quit date and who had been allocated nicotine or placebo patches but who reported any smoking in the previous week. A second sample of data collected at 8-weeks after the quit date from women who had smoked in the previous week were used as a sensitivity analysis. Not all women that had cotinine measured at 2-weeks returned for the 8-week visit, and 8-week data also included women who did not return at 2-weeks. We selected women from 2-weeks after the quit date rather than 8-weeks after the quit date for the main analysis, as this time point was earlier in gestation, and so nicotine metabolism changes since the baseline visit would likely be small and have less impact on findings (64).

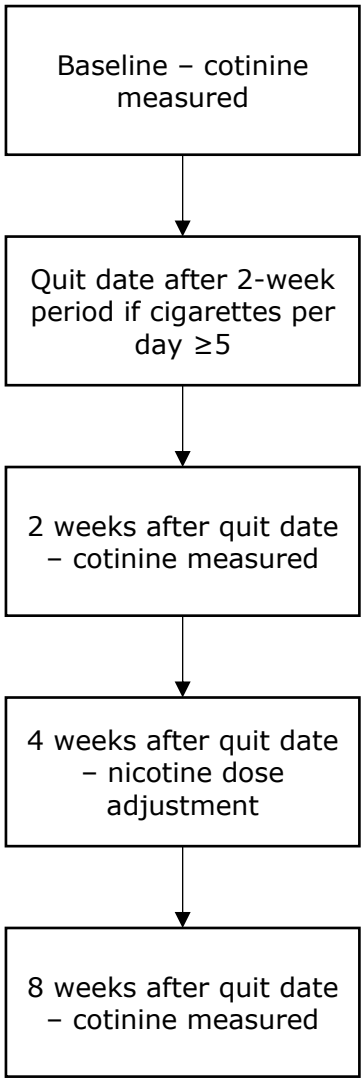

Figure 27 Flow chart to show each planned visit in the 'Study of Nicotine Patch in Pregnancy' relevant to the current study.

7.4.4 Analyses

For baseline data, continuous measures were reported as means with standard deviations (SD), and categorical measures were reported using frequencies and percentages. Participant and partner's smoking in the previous week were divided by seven, to achieve cigarettes smoked per day. T-tests were used to assess whether there were any systematic differences in baseline characteristics between women who were included and those

excluded from this study. We used a natural log transformation of salivary cotinine concentrations to achieve a normal distribution.

For both nicotine and placebo patch groups we used paired t-tests to assess 'within-participant' differences between cotinine, carbon monoxide, cravings, number of cigarettes smoked by the participant, and number of cigarettes smoked by their partner, measured at baseline and at 2-weeks. The same analyses were conducted using data from 8-weeks. For saliva cotinine, we present the back-transformed estimates of treatment differences, which is the ratio of the geometric means. Next, we used linear regression analysis to test for an interaction between the measures mentioned above and nicotine patch assignment. We then performed an exploratory analysis to identify whether the interactions were significant at increasing increments of baseline values in cotinine, carbon monoxide, cravings, number of cigarettes smoked by the participant, and number of cigarettes smoked by their partner. Findings are presented graphically. P-values less than 0.05 were deemed statistically significant. All analyses were conducted using STATA 15.

After undertaking the planned analyses, we generated a Bayes factor from the difference in saliva cotinine, using an online calculator (175). Bayes factors enable differentiation between whether there is no evidence of an effect, or whether it can be concluded that there is no effect. We were unable to identify any studies that investigated nicotine intake of concurrent smokers and NRT users in pregnancy, so an expected difference of 139.3 µg/L was taken from a study of nicotine intake outside of pregnancy (176). We used a conservative approach for estimation using a half-normal distribution, where the standard deviation is equal to the expected effect size.

7.5 Results

7.5.1 Descriptive statistics

In the SNIPP trial, 203 women were assigned to the nicotine patch arm and 199 women were assigned to the placebo patch arm. At 2-weeks after the quit-date, 167 (82.3%) and 148 (74.4%) women returned for the visit in the nicotine patch and placebo patch arms, respectively. In the nicotine patch arm, 149 (73.4%) had smoked in the week prior to the visit and 18 (8.9%) were abstinent, whereas, in the placebo group 131 (65.8%) had smoked in the week prior to the visit and 17 (8.5%) were abstinent. Overall, 12 women had missing cotinine data at this point and were excluded from the study, leaving a sample of 268 for analysis (146 in the nicotine group and 122 in the placebo group).

When comparing SNIPP trial participants excluded from this study with those included, it was found that more women in this study had a partner that smoked. **Table 3** gives baseline characteristics of women in both study groups and, using these descriptors, both groups were broadly similar. From the participants who provided 2-week data, those assigned nicotine patch had a mean age of 30 years and gestational age at baseline of 12.8 weeks; therefore, their mean gestational age at 2 weeks post quit date would be between 16 and 17 weeks.

Table 3 Participant baseline characteristics. n (%) or mean (standard deviation)

Characteristic	**Women on Nicotine Patch (n=146)**	**Women on Placebo Patch (n=122)**
Age (years)	29.70 (6.00)	28.88 (5.03)
BMI (kg/m^2)	25.52 (5.40)	25.21 (5.33)
Gestational age at baseline (weeks)	12.75 (3.24)	12.59 (5.42)
Ethnicity		
European	139 (95)	115 (94)
African	4 (3)	4 (3)
Asian	1 (1)	1 (1)
Other	2 (1)	2 (2)
Current cigarettes smoking per day		
5-10	66 (45)	55 (45)
11-20	69 (47)	50 (41)
21-30	7 (5)	16 (13)
>30	4 (3)	1 (1)
Fagerström Test for Cigarette Dependence[1]		
Very Low	32 (22)	20 (16)
Low	34 (23)	42 (34)
Medium	29 (20)	18 (15)
High	43 (29)	33 (27)
Very High	8 (6)	9 (7)
Partner smoking		
Yes	99 (69)	90 (75)
Saliva cotinine (ng/ml)	143.86 (82.81)	144.36 (74.33)
Expired air carbon monoxide (ppm)	11.81 (6.70)	12.22 (7.33)
French Tobacco Craving Questionnaire score	33.64 (8.60)	35.55 (9.53)

[1] FTCD is a 6-item test where answers are summed to yield a total score of 0-10. The higher the total score, the more intense is the patient's physical dependence on cigarettes. I.e. A score between 0-2 indicates a very low level of dependence on cigarettes, and 8-10 indicates a very high-level dependence on cigarettes (23).

7.5.2 Comparison of indicators of smoking intensity

Table 4 compares indicators of smoking intensity between baseline and 2 weeks after the quit date for pregnant smokers in both the placebo and nicotine patch groups. In the nicotine group, there was no significant difference between cotinine concentrations (ratio of geometric means = 0.94ng/ml, 95% CI's 0.83 to 1.07ng/ml; p=0.37, Bayes Factor=0.15), but CO concentrations significantly decreased from baseline to 2-weeks after the quit date (mean difference -3.03ppm, 95% CI's -4.17 to -1.89ppm; p<0.001). Whereas the placebo group exhibited a significant reduction in cotinine (ratio of geometric means = 0.68ng/ml, 95% CI's 0.59 to 0.78ng/ml; p<0.001) as well as a reduction in CO concentration (mean difference -2.02ppm, 95% CI's -3.81 to -0.22ppm, p<0.028). There were also significantly lower levels of craving, lower numbers of cigarettes smoked in the previous week and women's partners were reported to have smoked fewer cigarettes in both nicotine and placebo patch groups.

Table 4 also reports results for interaction tests between the indicators of smoking intensity and nicotine patch assignment. There was a significant interaction between nicotine patch assignment and a reduction in number of cigarettes smoked (p=0.046). This means that women assigned nicotine patches smoked less at week-2 compared to women assigned placebo patches. Interactions between the remaining indicators of smoking intensity and nicotine patch assignment were not significant. Upon further exploration it was discovered that there was an interaction between nicotine patch assignment and women with higher baseline cotinine concentrations (**Figure 28**). Women assigned nicotine patches with baseline saliva cotinine concentrations of approximately 90ng/ml and above had higher cotinine concentrations at week-2 compared to women assigned placebo patches.

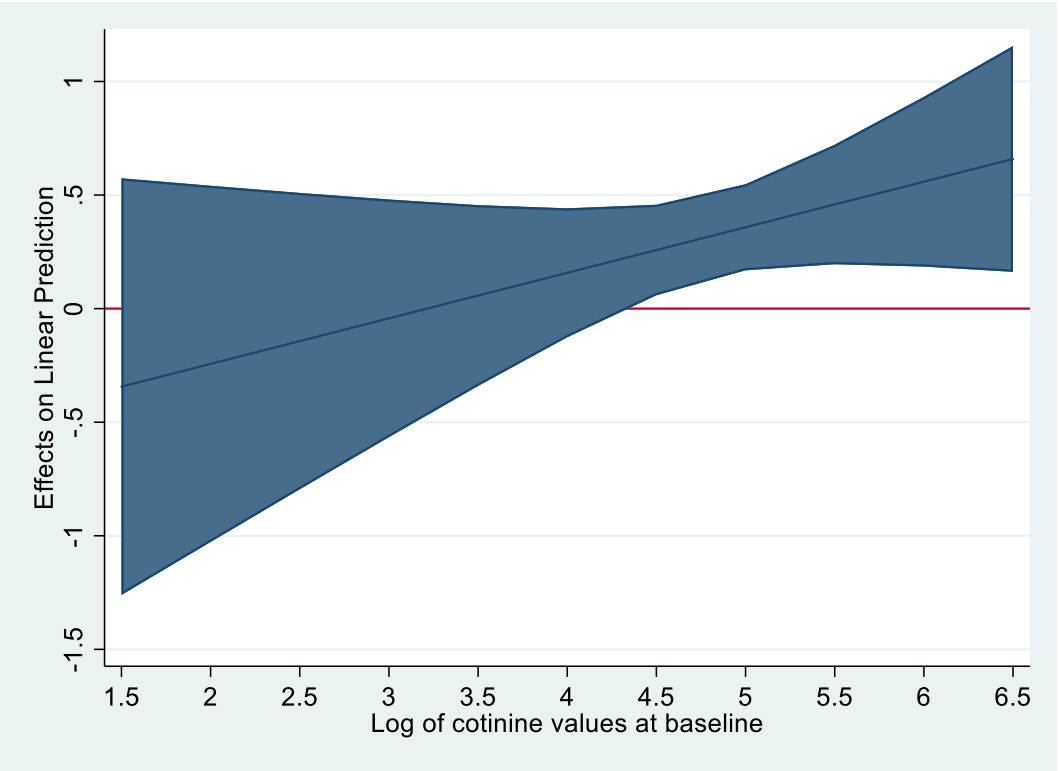

Figure 28 Graph to show interaction of nicotine patches on cotinine concentrations at 2-weeks with increasing baseline cotinine concentrations. The shaded area represents the 95% confidence intervals. As the shaded area for log cotinine >4.5 is above 0, there is a significant interaction of nicotine patches for an increase in cotinine at 2-weeks in women with log cotinine concentrations of greater than 4.5 (back-transformed to 90ng/ml), compared with placebo.

Table 4 Baseline to 2-weeks after the quit date 'within-participant' differences in indicators of smoking intensity in pregnant smokers by treatment group, with a significance test for interaction with nicotine patch.

	Nicotine Patch (n=146)				Placebo Patch (n=122)				
Characteristic	Baseline mean (SD)	2-weeks after quit date mean (SD)	Mean difference (95% CI)	p-value[1]	Baseline mean (SD)	2-weeks after quit date mean (SD)	Mean difference (95% CI)	p-value[2]	Interaction p-value[3]
Saliva cotinine[ŧ] (ng/ml)	117.83	111.14	0.94 (0.83 to 1.07)	0.370	122.46	83.01	0.68 (0.59 to 0.78)	<0.001	0.148
Expired air carbon monoxide (ppm)	11.77 (6.74)	8.74 (6.47)	-3.03 (-4.17 to -1.89)	<0.001	12.22 (7.33)	10.20 (9.08)	-2.02 (-3.81 to -0.22)	0.028	0.498
FTCQ-12[4]	33.75 (8.63)	31.38 (8.06)	-2.38 (-3.88 to -0.87)	0.002	35.84 (9.00)	33.36 (8.57)	-2.49 (-4.37 to -0.60)	0.010	0.317
Number of cigarettes smoked per day	12 (6)	6 (5)	-6 (-7 to -5)	<0.001	12 (6)	6 (6)	-6 (-7 to -5)	<0.001	0.046
Number of cigarettes partner smoked per day	17 (9)	15 (7)	-1 (-2 to 0)	0.026	16 (7)	14 (7)	-2 (-3 to -1)	0.003	0.168

Paired t-tests were used to compare differences at baseline and 2-weeks after the quit date. A linear model was used to test for an interaction of nicotine patch between baseline and 2-weeks.

[1]P-value for the difference between indicators of smoking intensity between baseline and 2-weeks, in the nicotine patch group

[2]P-value for the difference between indicators of smoking intensity between baseline and 2-weeks, in the placebo patch group

[3]P-value for interaction of nicotine patch with indicators of smoking intensity at baseline compared with at 2-weeks after the quit date

[4]FTCQ -12– French Tobacco Craving Questionnaire score

[ŧ]Back transformed saliva cotinine data. Means represent geometric means. Mean difference presented as ratio of geometric means.

7.5.3 Sensitivity analysis

In the sensitivity analysis, the 8-week data showed a similar pattern to the 2-week data (**Table 5**). There was no significant difference between cotinine concentrations at baseline and 8-weeks in the nicotine patch group (ratio of geometric means = 0.85ng/ml, 95% CI's 0.71 to 1.00ng/ml; p=0.055, Bayes Factor=0.12), however there were significant reductions for all other indicators of smoking intensity aside from craving score (mean difference = -1.69, 95% CI's -3.58 to 0.20 p=0.079). In women assigned placebo patches, there were significant reductions for all indicators of smoking intensity aside from expired CO concentration (mean difference -2.38ppm, 95% CI's -5.03 to 0.27ppm, p<0.077). The interaction tests found no significant interaction for nicotine patch assignment, however graphical exploration found that there was a significant interaction for nicotine patch assignment and participants that reported smoking between 100-250 cigarettes a week at baseline (**Figure 29**); in these women, assignment to nicotine patch was associated with having smoked fewer cigarettes in the previous seven days.

Table 5 Baseline to 8-weeks after the quit date 'within-participant' differences in indicators of smoking intensity in pregnant smokers by treatment group, with a significance test for interaction with nicotine patch.

	Nicotine Patch (n=86)				Placebo Patch (n=69)				
Characteristic	Baseline mean (SD)	8-weeks after quit date mean (SD)	Mean difference (95% CI)	p-value[1]	Baseline mean (SD)	8-weeks after quit date mean (SD)	Mean difference (95% CI)	p-value[2]	Interaction p-value[3]
Saliva cotinine[ŧ] (ng/ml)	116.35	98.24	0.85 (0.71 to 1.00)	0.055	118.13	87.11	0.74 (0.61 to 0.89)	0.002	0.874
Expired air carbon monoxide (ppm)	11.1 (6.3)	7.8 (5.7)	-3.3 (-4.6 to -1.9)	<0.001	12.5 (7.8)	10.1 (10.7)	-2.4 (-5.0 to 0.3)	0.077	0.844
FTCQ-12[4]	32.60 (8.30)	30.90 (7.56)	-1.69 (-3.58 to 0.20)	0.079	35.16 (7.99)	31.59 (7.02)	-3.57 (-5.65 to -1.48)	0.001	0.623
Number of cigarettes smoked per day	12 (8)	6 (4)	-7 (-8 to -5)	<0.001	12 (6)	7 (6)	-5 (-6 to -3)	<0.001	0.132
Number of cigarettes partner smoked per day	16 (7)	14 (6)	-2 (-3 to -1)	0.001	15 (7)	14 (7)	-2 (-3 to 0)	0.039	0.671

Paired t-tests were used to compare differences at baseline and 8-weeks after the quit date. A linear model was used to test for an interaction of nicotine patch between baseline and 8-weeks.

[1]P-value for the difference between indicators of smoking intensity between baseline and 8-weeks, in the nicotine patch group

[2]P-value for the difference between indicators of smoking intensity between baseline and 8-weeks, in the placebo patch group

[3]P-value for interaction of nicotine patch with indicators of smoking intensity at baseline compared with at 8-weeks after the quit date

[4]FTCQ -12– French Tobacco Craving Questionnaire score

[ŧ]Back transformed saliva cotinine data. Means represent geometric means. Mean difference presented as ratio of geometric means.

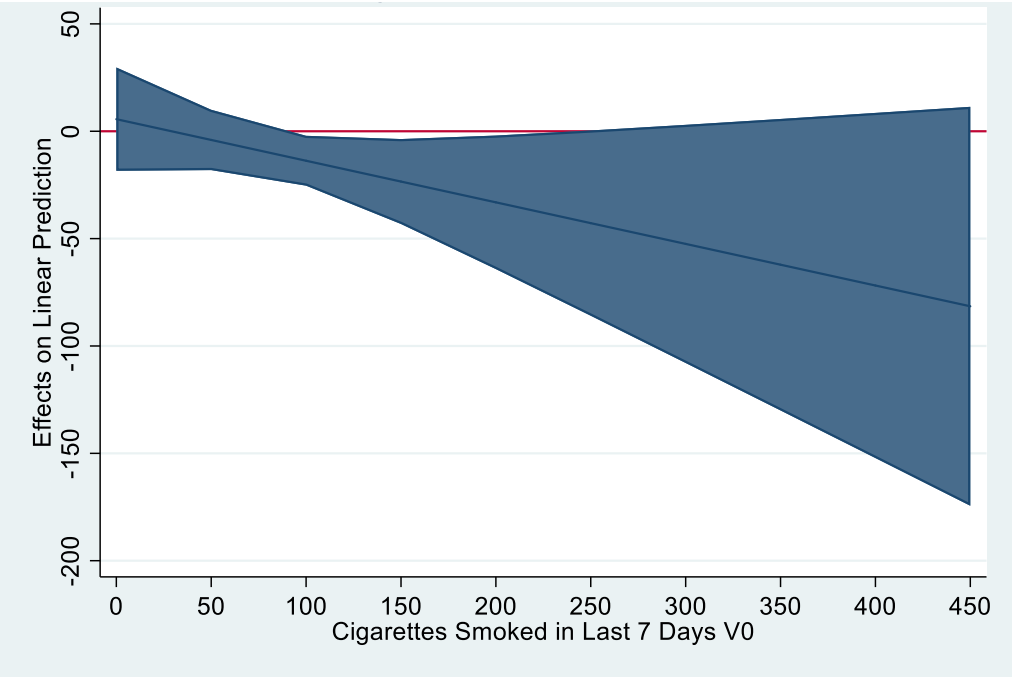

Figure 29 Graph to show interaction of nicotine patches on cigarettes smoked at 2-weeks with increasing number of cigarettes smoked at baseline. The shaded area represents the 95% confidence intervals. As the shaded area for number of cigarettes smoked between 100-250, is below 0, there is a significant interaction of nicotine patches for a reduction of cigarettes smoked at 8-weeks in women that smoked between 100-250 cigarettes in the week prior to baseline compared with placebo.

7.6 Discussion

7.6.1 Key findings

Our findings show that, women prescribed nicotine patches but also admitted smoking had similar cotinine concentrations to those generated when they only smoked. These women also reported smoking less and had lower expired air carbon monoxide readings than when they smoked prior to their quit attempt. In comparison, smokers issued with placebo patches had lower cotinine concentrations than when smoking, they also showed reductions in numbers of cigarettes smoked and expired CO concentrations. Our results also indicate that women who smoke and use nicotine patches, smoke less later in pregnancy.

7.6.2 Strengths and weaknesses

A limitation to our study is that, whilst we know that women included in this study were prescribed nicotine patches, we have very limited information about how much they used these. However, as study measurements at 2- and 8-week follow up were taken with the intention of personalising the nicotine doses which women received from patches, it seems very likely that women who attended these appointments were still using these. Furthermore, the SNIPP trial also reports (where adherence data exists) the median self-reported adherence rate was 85% (51).

Another possible limitation concerns the validity of women's reports of smoking or not smoking in the week prior to having 2- and 8-week measurements taken. In SNIPP, women were defined as smokers if they had reported any smoking in the week prior to a study visit and this was validated by an expired CO reading. However, expired air CO can only reliably validate smoking status over the previous 6 hours and (177), although some women may have over-or under estimated the number of

cigarettes smoked in the previous week, we could only accurately quantify tobacco smoke exposure in the 6 hours prior to CO measurement. Nevertheless, this could only have had a major impact on findings if women generally under-reported their smoking in the week prior to follow up appointments and, in the 6 hours before follow-up appointments tried to smoke less than they had reported they were doing. It seems unlikely that trial participants would do this before attending a nicotine patch dose-titration appointment.

A strength of this study is that the data were obtained as part of a well-conducted randomized controlled trial and included reported smoking behaviour with concurrent CO and cotinine estimation at several time points. To the authors' knowledge, there has been no previous study that has investigated smoking behaviour and CO exposure from concurrent use of nicotine patches and smoking in pregnancy. Hence, we believe it makes an original contribution to the field. Another strength is that comparisons are based on 'within-participant' measurements; this means that inter-participant variations are very unlikely to explain study findings. Indeed, with this study design one would only expect findings to be affected by characteristics of women which were prone to change between baseline and follow up. Women's nicotine metabolic rates (NMR) increase as pregnancy progresses and these would be expected to affect their plasma nicotine concentrations and so potentially their cravings and intensity of smoking too (64, 69). However, any effect would seem to be marginal as, even in the placebo group, women reported smoking fewer cigarettes. Also, as pregnancy-related NMR (nicotine metabolic ratio) acceleration is generally complete by the end of the first trimester and women's mean gestation at baseline was ~ 13 weeks, there may have been little scope for this factor to have any influence. It seems likely, therefore that differences reported

reflect differences in smoking behaviour and not changes in women's physiology during pregnancy.

7.6.3 Discussion in context of previous literature

Our study informs about cotinine concentrations in pregnant women who use nicotine patches but are not abstinent from smoking. Our findings show that cotinine concentrations in such women were no higher than when they were smoking. Additionally, women included in this study had simultaneous and statistically significant reductions in their cigarette use, validated by a reduction in expired carbon monoxide. This suggests that when pregnant women use nicotine patches and smoke, they smoke less than they would without if they were not using nicotine patches. This is important, as it could influence how women are advised to use NRT in pregnancy, i.e. encouraged to continue using NRT despite a relapse.

We are unaware of any previous studies measuring cotinine or CO in smokers who concurrently use NRT during pregnancy. A systematic review and meta-analysis that aimed to identify and describe studies which report nicotine or cotinine concentrations in pregnant women when smoking and subsequently when abstinent from smoking and using NRT, concluded that amongst pregnant women who quit smoking, standard-dose NRT generates lower nicotine exposure than smoking (161). The meta-analysis compared cotinine exposures when pregnant women smoke with those when they use NRT and found that concentrations were on average 75.3 ng/ml lower when abstinent and using NRT than when the same women smoked (161). In SNIPP, salivary cotinine concentrations at baseline (when smoking) were compared to cotinine concentrations at 1 month in women that had stopped smoking but were using nicotine patches. Cotinine concentrations were 98.5ng/ml while smoking, but only 62.8ng/ml while using nicotine patches (51). In our study we found that women that were assigned the placebo

patch but also admitted to smoking, also exhibited reduced cotinine concentrations compared to those when smoking alone.

Most studies in the above review, used lower nicotine doses than were used by participants in this manuscripts' analyses; other than SNIPP, studies used standard rather than higher doses of nicotine and these delivered no more than 15mg cotinine in 16 hours or the 24-hour equivalent (161). Thus, when pregnant smokers become abstinent and adhere with such 'standard' doses of NRT, they are on average exposed to less nicotine than from smoking (161). In SNIPP, patch doses were adjusted according to the previous saliva cotinine determination to optimize the nicotine substitution leading to somewhat higher mean nicotine doses than usual (18 mg/day, SD=6.8). It is expected that the dose adjustment would improve nicotine substitution, thus it is possible that women assigned nicotine patches in the 8-week sample would have higher cotinine concentrations than they did at baseline. Despite this adjustment, there was no significant difference in cotinine concentrations in women that were assigned nicotine patches and admitted to smoking to those when smoking alone. This also suggests that smoking and using nicotine patches of 'standard' doses, may lead to lower cotinine concentrations during pregnancy than smoking alone, prior to pregnancy.

Our findings provide the first data we are aware of which quantifies pregnant women's smoking behaviour when using nicotine patches and this suggests that when pregnant women use nicotine patches as part of a quit attempt, but they also smoke, they smoke less than they did before the quit attempt started. This means that their exposure to the toxic products of burnt tobacco is reduced. A possible reason for this is that women who continue to smoke when using nicotine patches obtain nicotine from both patches and tobacco and nicotine delivered from patches reduces women's cravings such that they feel less need to 'top up' concentrations of nicotine in their

body fluids through smoking. This suggests that clinicians can reassure women that it is ok to smoke and use nicotine patches if, ultimately, they are trying for abstinence.

7.6.4 Conclusions

In conclusion, despite having similar cotinine exposure to that from cigarette smoking, pregnant women who use nicotine patches and smoke, smoke less and exhale less CO, so their exposure to other tobacco smoke toxins is likely to be lower too.

Chapter 8: Conclusions and recommendations

8.1 Introduction

The overall aim is to raise hypotheses regarding ways in which NRT use in pregnancy might be changed such that it has greater potential to be effective. This aim was achieved through investigating the efficacy, safety and impacts on smoking intensity of Nicotine Replacement Therapy used for smoking cessation in pregnancy, and was facilitated by the use of TSA.

This chapter summarises results from this thesis in context to each of the objectives detailed in **Chapter 2**. This chapter also describes how the results have been disseminated, their potential implications for policy and practice, and provides suggestions for possibilities of future research that could be undertaken relating to the use of NRT for smoking cessation in pregnancy. Re-prints of published papers from this thesis are included in **Appendix B**, and a full list of courses attended, training, and publications produced from and during the writing of this thesis are detailed in **Appendix C**.

8.2 Summary of thesis findings

8.2.1 Objective I: To use conventional systematic review and meta-analysis to determine the efficacy and safety of NRT used during pregnancy for smoking cessation in later pregnancy and after childbirth.

A systematic review with meta-analysis was conducted to assess the efficacy and safety of NRT for smoking cessation in pregnancy. There was low certainty evidence that NRT used alongside usual care may increase smoking abstinence in later pregnancy. However, subgroup analysis of the more robust placebo-controlled trials suggest that there may not be an effect of NRT on smoking abstinence. Evidence was inconsistent regarding

the evidence of NRT having either a positive or negative impact on birth outcomes.

This systematic review was conducted as part of a broader Cochrane review assessing the efficacy of pharmacological interventions, including bupropion, for smoking cessation in pregnancy. This review has been published and is available from the *Cochrane Library* (178). The results of this review have been discussed at the NICE Public Health Advisory Committee tobacco meeting for preventing uptake, promoting quitting and treating dependence. The purpose of this meeting was to utilise findings from the review to update the current NICE guidelines for stopping smoking in pregnancy.

8.2.2 Objective II: To describe the limitations of meta-analysis and demonstrate how trial sequential analysis methodology can be used to supplement the findings of meta-analysis.

This thesis provides an overview of how TSA can be used alongside meta-analysis to assess whether there is sufficient evidence to conclude a clinically important treatment effect, no evidence of an effect, or absence of evidence. Using example outputs from TSA, this thesis demonstrates how TSA can be interpreted, and how they are affected when new studies are included. As a worked example, TSA is then applied to a systematic review of NRT for smoking cessation in pregnancy (**Objective III**).

Whilst TSA methodology and software has been described in depth previously (93), and the methodology has been utilised increasingly in recent years (179, 180), this thesis attempts to explain TSA in a way that is accessible to a broader range of researchers.

8.2.3 Objective III: To determine whether there is sufficient information in the meta-analyses presented for objective I above regarding the efficacy and safety of NRT for smoking cessation in later pregnancy.

Using the meta-analyses conducted in the systematic review for **Objective I**, TSA was applied to the primary outcomes of biochemically validated smoking cessation at the latest point in pregnancy; mean birthweight; and miscarriage and spontaneous abortion. TSA methods were able to ascertain whether there was sufficient information in the meta-analyses regarding the efficacy and safety of NRT for smoking cessation in later pregnancy.

According to TSA, over 8,000 participants to placebo-controlled trials may be needed in order to arrive at a firm conclusion regarding the efficacy of NRT for smoking cessation in pregnancy. Additionally, an excess of 9,000 and 30,000 participants respectively would be needed for a strong conclusion surrounding the positive or negative impact of NRT on mean birthweight; and miscarriage and spontaneous abortion.

8.2.4 Objective IV: To demonstrate how trial sequential analysis can alternatively be utilised to calculate trial sample size, using results from feasibility and pilot studies.

Using data from feasibility and pilot RCT's testing MiQuit, a text message-based smoking cessation intervention for pregnant women, TSA was used to estimate the sample size of future RCTs for a more conclusive decision regarding intervention benefit. The TSA estimated sample size required just over half the participants than that calculated by the traditional sample size calculation methodology.

Prior to this study, no paper has described utilising TSA in this way. This relatively simple use of feasibility and pilot trial data with TSA, could save researchers significant resources, thus leading to more efficient utilisation of funds. A paper describing this methodology has been submitted to *BMC Medical Research Methodology*. This paper has been peer reviewed and a response alongside amendments has been submitted, and I now await a decision. A pre-print of this paper is available on *Research Square* (119).

8.2.5 Objective V: To use the SNIPP trial to investigate the differences in indicators of smoking intensity in pregnant women when smoking before using NRT, and when using NRT and smoking concurrently.

Using a cohort of participants from the SNIPP trial, within-participant differences in indicators of smoking intensity were compared in women who both smoked and used nicotine or placebo patches. Women who both smoked and used nicotine patches concurrently had similar cotinine concentrations as those when they were only smoking. These women also reported smoking fewer cigarettes when using nicotine patches, ultimately reducing their exposure to toxins in tobacco smoke.

Prior to this research, no other studies had measured pregnant women's smoking behaviour when using nicotine patches as part of a quit attempt. This is mainly because in practice, pregnant women are usually advised to stop using NRT if they relapse to smoking. The SNIPP trial was different, as it allowed women to continue using NRT despite relapse. This research was shared through publication in *Addiction Journal* (162).

8.3 Policy and practice implications

The most recent UK guidelines and recommendations for stopping smoking in pregnancy and in the first year of childbirth were published in 2010 (181). These guidelines were reviewed in 2015, and are currently being updated. Research from this thesis adds to the current knowledge on smoking in pregnancy, and will inform the next update of guidelines.

The current recommendations surrounding the use of NRT for smoking cessation during pregnancy are based on 'mixed evidence on the effectiveness of NRT in helping women to stop smoking during pregnancy' (181). The evidence from the systematic review performed in **Chapter 5** suggests that NRT may be effective for smoking cessation in pregnancy, however there is uncertainty surrounding this evidence. It is also unclear whether NRT affects the risk of adverse pregnancy and infant outcomes, but there is no evidence that it is harmful. One study suggests that NRT improves child development outcomes at two years (156). Nevertheless, the 2010 guidelines state that NHS Stop Smoking Services should only suggest using NRT if smoking cessation attempted without this first (181). Findings from the systematic review, conducted as part of a comprehensive Cochrane Review, are to be used in the latest update of NICE guidelines. In late 2019, I was invited and attended the NICE Public Health Advisory Committee tobacco meeting for preventing uptake, promoting quitting and treating dependence. Here, the Cochrane review was discussed in detail in the context of updating the guidelines.

Current NICE guidelines state that women should only be prescribed NRT once they have stopped smoking, and should only be prescribed two weeks of NRT from an agreed quit-date (181). Furthermore, subsequent prescriptions of NRT should only be provided when women have

demonstrated that they are still abstinent (181). This cautious approach to prescribing NRT is to prevent a possible increase of nicotine intake, in case women continue to smoke whilst using NRT. Findings from **Chapter 7** provide evidence to suggest that women who smoke and use NRT concurrently do not have higher nicotine concentrations compared to when they are only smoking. These women also smoked fewer cigarettes. These findings could allay some fears regarding higher nicotine concentrations with continued use of NRT despite relapse, and suggest that policy could be adapted such that, as long as women continue trying to stop smoking, if they have a brief relapse to smoking, they should continue using NRT.

This thesis has provided some potential recommendations for updates of policy and practice for the use of NRT for smoking cessation in pregnancy however, further research is also required. Possibilities for future research are detailed in the following section.

8.4 Possibilities for further research

TSA appears to be growing in popularity and there is a developing knowledge base of NRT use during pregnancy. The research summarised in **Section 8.2** adds to the current understanding on this subject. However, this thesis has opened up potential avenues for future research and these are explored in the following section.

In **Chapter 5** a systematic review was performed to determine the efficacy and safety of NRT used during pregnancy for smoking cessation in pregnancy. Although overall evidence pointed towards a positive effect of NRT on smoking abstinence in pregnancy, a subgroup of low-bias trials suggests the effect is smaller or that there may be no effect at all. These findings coupled with results from the TSA in **Chapter 6,** suggest that further placebo-controlled trials are needed, for greater certainty regarding

the efficacy and safety of NRT for smoking cessation in pregnancy. The TSA performed in **Chapter 6** recommended that a further 8,500 participants would need to be recruited in placebo-controlled trials. This would require substantial time and resources, and may not be practical over a short period of time.

Instead, focus could be spent understanding why NRT apparently works less well in pregnancy than it does in the general population. For example, the systematic review in **Chapter 5** found that patch adherence was low. Future research should seek to understand why this is and improve it and use an appropriate behavioural strategy to enhance adherence in future trials of NRT. Additionally, the majority of studies in the systematic review in **Chapter 5** used 15mg nicotine patches, whereas in the general population there is evidence that 25 mg/16-hour patches are more effective than 15 mg/16-hour patches (164). Consequently, trials are needed in pregnant women using either higher-dose nicotine patches or combination of patch plus rapid-acting forms of NRT, which are also more effective (164). As of yet, there are no complete trials of electronic cigarettes in pregnancy. In a non-pregnant population, a recent living systematic review found that electronic cigarettes are not only effective for quitting smoking, but they are more effective than NRT (46). There is currently one ongoing trial of electronic cigarettes for smoking cessation in pregnancy, however further research should be conducted.

With regards to safety, **Chapter 5** found no consistent evidence of NRT having either a positive or adverse impact on birth outcomes. Coupled with the TSA performed in **Chapter 6**, nearly 10,000 participants would need to be recruited for greater certainty over the effects of NRT on mean birthweight, and in excess of 30,000 participants would need to be recruited for greater certainty over the effects of NRT on miscarriage and spontaneous

abortion. As detailed above, recruiting this number of participants is highly unlikely, thus alternative methodology may be required. A possible way to investigate this further, would be to perform an individual patient data meta-analysis, by combining participants with reported NRT use and outcome data from all NRT trials in pregnancy.

The study in **Chapter 7** provides data quantifying concurrent smoking and NRT use during pregnancy and as it is the first study of its kind in pregnancy, the findings are important. However, this study was limited by insufficient data regarding adherence to patches. Future trials should aim to collect comprehensive adherence data; where possible, this will facilitate a more accurate analysis to see whether greater adherence to NRT leads to fewer cigarettes smoked.

In previous studies investigating the efficacy and safety of NRT during pregnancy, women were told not to use NRT if they had relapsed to smoking. Findings from the study in **Chapter 7** suggest that in future trials of NRT in pregnancy, women should be encouraged to continue using NRT despite relapse if their ultimate goal is still abstinence. Indeed, safety is of paramount importance and so if a decline in cigarette consumption is not observed, prescription of NRT should be reviewed in those individual cases. The latest RCT of nicotine inhaler in pregnancy advised participants to continue the use of the inhaler as long as they were actively trying to quit smoking. Although the inhaler group did not have a higher quit rate than the placebo group, they did have significantly decreased risks of delivering a preterm or a low birth weight infant (151). Data permitting, a secondary analysis of this trial similar to that conducted in **Chapter 7** could help corroborate the findings.

In **Chapter 4**, an alternate use for TSA, to estimate the sample size(s) for future trial(s) based on pilot and feasibility trial data, is presented. This method could be utilised in the future in trials of NRT in pregnancy. Running large scale trials is expensive and resource intensive. Where funding is highly competitive or limited, it may be beneficial to use TSA to estimate an alternative sample size. For example, if feasibility and pilot studies investigating a combination of nicotine patch and fast-acting forms of NRT in pregnancy are successful, using TSA to estimate the sample size for a more definitive trial could save on costs, which could be more attractive to potential funders. If the resulting definitive trial is unsuccessful, recruiting fewer participants means that excess resources were not wasted on a much larger trial if a traditional sample size calculation was performed.

Chapter 3 describes the more traditional use of TSA, and as mentioned above, **Chapter 6** applies TSA to the meta-analysis in **Chapter 5**. The more traditional use of TSA has grown in popularity in recent years and has been used in a number of systematic reviews (179, 180, 182), including smoking cessation reviews (183, 184). Despite the limitations of TSA, discussed in **Section 3.5.5**, as long as it is used appropriately, i.e. planned prospectively with the assumptions underlying the design well planned and justified, then TSA should continue to be used to provide additional context to a meta-analysis.

8.5 Overall conclusions

This thesis has demonstrated two possible uses of TSA in the context of smoking cessation in pregnancy, and has also provided evidence to illustrate some ways in which the use of NRT in pregnancy might be changed, such that it has greater potential to be found effective, as it is in non-pregnant smokers. NRT used for smoking cessation in pregnancy may increase

smoking cessation rates in late pregnancy. According to TSA, there is uncertainty regarding the efficacy of NRT use for smoking cessation during pregnancy compared to control, and further placebo-controlled trials are needed to arrive at a firm conclusion. Although TSA suggests more research is required for a firm conclusion, the general trend appears that NRT as it has previously been trialled, may not be effective for smoking cessation in pregnant women. Further trials should focus on what can be done differently in future. Following successful feasibility and pilot trials, these future trials could make use of TSA for sample size estimation, to reduce costs and resources required. When pregnant women use nicotine patches as part of a quit attempt, but they also smoke, they smoke less than they did before the quit attempt started. This means that their exposure to the toxic products of burnt tobacco is reduced. Despite having similar cotinine exposure to that from cigarette smoking, pregnant women who use nicotine patches and smoke, smoke less and exhale less CO, so their exposure to other tobacco smoke toxins is likely to be lower too. Overall this thesis should encourage further investigation of new techniques to trial NRT in pregnant women, and reassure policy makers to encourage NRT use despite relapse.

8.6 Personal development and development of research skills

Throughout the duration of this PhD I have improved and developed my research skills, and I have learned a number of new methods and techniques that will benefit me in the future as an independent researcher. Attending the 'Joanna Briggs Institute (JBI) Comprehensive Systematic Review Training Program', further increased my ability and efficiency to locate journal articles and other relevant resources to include throughout the

thesis. The JBI program, alongside the two Master's in Public Health (MPH) statistics modules I undertook, gave me a solid foundation for much of the statistical analysis performed in the thesis. The analysis in **Chapter 7** was performed using STATA, and involved performing t-tests and regression analysis – which these modules were particularly useful for. Furthermore, lectures and practical sessions on meta-analysis and forest-plots in both the MPH modules and the JBI program, provided me the basis to better understand trial sequential analysis methodology. A large part of this thesis discusses and utilises TSA methodology and software, which was an entirely new concept to me prior to this PhD. I was able to develop a thorough understanding of TSA through independent research of literature, with support from my PhD supervisors and external collaboration.

During the second year of my PhD I spent 6 months in Paris, France on a placement to work alongside my supervisor, Ivan Berlin. This enabled me to experience a research environment outside of the UK. Through this experience I was able to twice present my TSA work to a French audience that had not previously heard of TSA. Through dissemination of my research via poster and oral presentations at both national and international conferences, as well as writing papers as a first author (**Appendix C**), I have been able to develop my capability as an independent researcher.

Additionally, during the fourth year of my PhD, I have been working as a Research Assistant on a National Institute for Health Research - School for Primary Care Research (NIHR SPCR) funded programme investigating infant and child primary and secondary health care costs associated with mode of childbirth and prematurity. As part of this project I have developed an understanding for costing of primary and secondary healthcare episodes. This project has also exposed me to the use and management of large datasets through Clinical Practice Research Datalink (CPRD) data. Finally,

as lead applicant, I was able to secure NIHR SPCR funding for a systematic review comparing nicotine concentrations generated by concurrent smoking and nicotine replacement therapy. This successful application has given me experience in writing grant applications, which will benefit me greatly for a future career in academia.

References

1. WHO. Tobacco (Fact Sheet): World Health Organization; 2020 [Available from: https://www.who.int/news-room/fact-sheets/detail/tobacco. (Accessed: 05-09-2020.)]

2. WHO. WHO Report on the Global Tobacco Epidemic. 2013. [Available from: https://www.who.int/tobacco/global_report/2013/en/. (Accessed: 29-03-2017.)]

3. Lange S, Probst C, Rehm J, Popova S. National, regional, and global prevalence of smoking during pregnancy in the general population: a systematic review and meta-analysis. The Lancet Global Health. 2018;6(7):e769-e76.

4. Al-Sahab B, Saqib M, Hauser G, Tamim H. Prevalence of smoking during pregnancy and associated risk factors among Canadian women: a national survey. BMC pregnancy and childbirth. 2010;10(1):24.

5. Lanting CI, Segaar D. Smoking during pregnancy: trends between 2001 and 2010. Nederlands tijdschrift voor geneeskunde. 2012;156(46):A5092-A.

6. Tappin DM, MacAskill S, Bauld L, Eadie D, Shipton D, Galbraith L. Smoking prevalence and smoking cessation services for pregnant women in Scotland. Substance abuse treatment, prevention, and policy. 2010;5(1):1.

7. Department of Health. Healthy Lives, Healthy People: A Tobacco Control Plan for England. 2011 11/03/11. [Available from: https://www.gov.uk/government/publications/the-tobacco-control-plan-for-england. (Accessed: 29-03-2017.)]

8. NHS Digital. Statistics on Women's Smoking Status at Time of Delivery 2017 [Available from: http://www.content.digital.nhs.uk/catalogue/PUB24222/stat-wome-smok-time-deli-eng-q4-16-17-rep.pdf. (Accessed: 30-03-2017.)]

9. Department of Health. Towards a Smokefree Generation. A Tobacco Control Plan for England. 2017 July 2017. [Available from: https://www.gov.uk/government/publications/towards-a-smoke-free-generation-tobacco-control-plan-for-england. (Accessed: 30-03-2017.)]

10. Office for National Statistics. Births in England and Wales: 2016 2016 [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/birthsummarytablesenglandandwales/2016#the-number-of-stillbirths-decreased-in-2016. (Accessed: 31-03-2017.)]

11. Flenady V, Wojcieszek AM, Middleton P, Ellwood D, Erwich JJ, Coory M, et al. Stillbirths: recall to action in high-income countries. The Lancet. 2016;387(10019):691-702.

12. Bjørnholt SM, Leite M, Albieri V, Kjaer SK, Jensen A. Maternal smoking during pregnancy and risk of stillbirth: results from a nationwide

Danish register-based cohort study. Acta Obstet Gynecol Scand. 2016;95(11):1305-12.

13. Pineles BL, Hsu S, Park E, Samet JM. Systematic Review and Meta-Analyses of Perinatal Death and Maternal Exposure to Tobacco Smoke During Pregnancy. American journal of epidemiology. 2016;184(2):87-97.

14. Marufu TC, Ahankari A, Coleman T, Lewis S. Maternal smoking and the risk of still birth: systematic review and meta-analysis. BMC Public Health. 2015;15(1):239.

15. Pineles BL, Park E, Samet JM. Systematic review and meta-analysis of miscarriage and maternal exposure to tobacco smoke during pregnancy. American journal of epidemiology. 2014;179(7):807-23.

16. Tobacco Advisory Group of the Royal College of Physicians. Passive smoking and children. London: Royal College of Physicians; 2010. [Available from: https://cdn.shopify.com/s/files/1/0924/4392/files/passive-smoking-and-children.pdf?15599436013786148553. (Accessed: 01-04-2017.)]

17. Jornayvaz FR, Vollenweider P, Bochud M, Mooser V, Waeber G, Marques-Vidal P. Low birth weight leads to obesity, diabetes and increased leptin levels in adults: the CoLaus study. Cardiovascular diabetology. 2016;15(1):73.

18. Cnattingius S. The epidemiology of smoking during pregnancy: smoking prevalence, maternal characteristics, and pregnancy outcomes. Nicotine & Tobacco Research. 2004;6(Suppl_2):S125-S40.

19. Kyrklund-Blomberg NB, Cnattingius S. Preterm birth and maternal smoking: risks related to gestational age and onset of delivery. American journal of obstetrics and gynecology. 1998;179(4):1051-5.

20. Kyrklund-Blomberg NB, Granath F, Cnattingius S. Maternal smoking and causes of very preterm birth. Acta Obstet Gynecol Scand. 2005;84(6):572-7.

21. Fleming P, Blair PS. Sudden infant death syndrome and parental smoking. Early Hum Dev. 2007;83(11):721-5.

22. Krous HF, Beckwith JB, Byard RW, Rognum TO, Bajanowski T, Corey T, et al. Sudden infant death syndrome and unclassified sudden infant deaths: a definitional and diagnostic approach. Pediatrics. 2004;114(1):234-8.

23. Hackshaw A, Rodeck C, Boniface S. Maternal smoking in pregnancy and birth defects: a systematic review based on 173 687 malformed cases and 11.7 million controls. Hum Reprod Update. 2011;17(5):589-604.

24. McCrory C, Layte R. Prenatal exposure to maternal smoking and childhood behavioural problems: a quasi-experimental approach. J Abnorm Child Psychol. 2012;40(8):1277-88.

25. Doll R, Peto R, Boreham J, Sutherland I. Mortality in relation to smoking: 50 years' observations on male British doctors. Bmj. 2004;328(7455):1519.

26. McAndrew F, Thompson J, Fellows L, Large A, Speed M, Renfrew MJ. Infant Feeding Survey 2010: Health and Social Care Information Centre. 2012. [Available from: https://digital.nhs.uk/data-and-information/publications/statistical/infant-feeding-survey/infant-feeding-survey-uk-2010. (Accessed: 03-04-2017.)]

27. Leonardi-Bee J, Jere ML, Britton J. Exposure to parental and sibling smoking and the risk of smoking uptake in childhood and adolescence: a systematic review and meta-analysis. Thorax. 2011;66(10):847-55.

28. Gray R, Bonellie SR, Chalmers J, Greer I, Jarvis S, Kurinczuk JJ, et al. Contribution of smoking during pregnancy to inequalities in stillbirth and infant death in Scotland 1994-2003: retrospective population based study using hospital maternity records. Bmj. 2009;339:b3754.

29. Moore M, for ASH and Smoking in Pregnancy Challenge Group. Smokefree Skills: An assessment of maternity workforce training. 2017 July 2017. [Available from: https://ash.org.uk/wp-content/uploads/2017/07/Smokefree-skills-an-assessment-of-maternity-workforce-training.pdf. (Accessed: 03-04-2017.)]

30. Harmer C, Memon A. Factors Associated With Smoking Relapse in the Postpartum Period: An Analysis of the Child Health Surveillance System Data in Southeast England. Nicotine & Tobacco Research. 2013;15(5):904-9.

31. Räisänen S, Sankilampi U, Gissler M, Kramer MR, Hakulinen-Viitanen T, Saari J, et al. Smoking cessation in the first trimester reduces most obstetric risks, but not the risks of major congenital anomalies and admission to neonatal care: a population-based cohort study of 1 164 953 singleton pregnancies in Finland. J Epidemiol Community Health. 2014;68(2):159-64.

32. Ferguson J, Bauld L, Chesterman J, Judge K. The English smoking treatment services: one-year outcomes. Addiction. 2005;100 Suppl 2:59-69.

33. Campbell KA, Cooper S, Fahy SJ, Bowker K, Leonardi-Bee J, McEwen A, et al. 'Opt-out' referrals after identifying pregnant smokers using exhaled air carbon monoxide: impact on engagement with smoking cessation support. Tobacco Control. 2017;26:300-6.

34. Brenner H, Mielck A. The role of childbirth in smoking cessation. Preventive medicine. 1993;22(2):225-36.

35. Wisborg K, Kesmodel U, Henriksen TB, Olsen SF, Secher NJ. Exposure to tobacco smoke in utero and the risk of stillbirth and death in the first year of life. American journal of epidemiology. 2001;154(4):322-7.

36. Slade P, Laxton-Kane M, Spiby H. Smoking in pregnancy: The role of the transtheoretical model and the mother's attachment to the fetus. Addict Behav. 2006;31(5):743-57.

37. Stella MY, Park CH, Schwalberg RH. Factors associated with smoking cessation among US pregnant women. Maternal and Child Health Journal. 2002;6(2):89-97.

38. Schneider S, Huy C, Schuetz J, Diehl K. Smoking cessation during pregnancy: a systematic literature review. Drug and alcohol review. 2010;29(1):81-90.

39. Torrent M, Sunyer J, Cullinan P, Basagaña X, Harris J, García O, et al. Smoking cessation and associated factors during pregnancy. Gaceta Sanitaria. 2004;18(3):184-9.

40. Chamberlain C, O'Mara-Eves A, Porter J, Coleman T, Perlen SM, Thomas J, et al. Psychosocial interventions for supporting women to stop smoking in pregnancy. Cochrane Database of Systematic Reviews. 2017(2).

41. Lumley J, Chamberlain C, Dowswell T, Oliver S, Oakley L, Watson L. Interventions for promoting smoking cessation during pregnancy. Cochrane Database Syst Rev. 2009(3):CD001055.

42. Cahill K, Lancaster T, Perera R. Pharmacological interventions for smoking cessation: an overview of reviews. Cochrane Database of Systematic Reviews. 2011;9.

43. Cahill K, Stead LF, Lancaster T. Nicotine receptor partial agonists for smoking cessation. Cochrane Database Syst Rev. 2012;4:CD006103.

44. Stotts AL, Northrup TF, Cinciripini PM, Minnix JA, Blalock JA, Mullen PD, et al. Randomized, Controlled Pilot Trial of Bupropion for Pregnant Smokers: Challenges and Future Directions. Am J Perinatol. 2015;32(04):351-6.

45. Rore C, Brace V, Danielian P, Williams D. Smoking cessation in pregnancy. Expert opinion on drug safety. 2008;7(6):727-37.

46. Hartmann-Boyce J, McRobbie H, Lindson N, Bullen C, Begh R, Theodoulou A, et al. Electronic cigarettes for smoking cessation. Cochrane Database of Systematic Reviews. 2020(10).

47. US Department of Health and Human Services. E-cigarette use among youth and young adults: a report of the Surgeon General. Atlanta, GA: US Department of Health and Human Services, CDC; 2016.

48. WHO. WHO report on the global tobacco epidemic 2019: Offer help to quit tobacco use. 2019. [Available from: https://www.who.int/tobacco/mpower/offer/en/index4.html. (Accessed: 12-08-2020.)]

49. Stead LF, Perera R, Bullen C, Mant D, Hartmann-Boyce J, Cahill K, et al. Nicotine replacement therapy for smoking cessation. Cochrane Database Syst Rev. 2012;11(11).

50. Coleman T, Chamberlain C, Davey MA, Cooper SE, Leonardi-Bee J. Pharmacological interventions for promoting smoking cessation during pregnancy. Cochrane Database Syst Rev. 2015(12):CD010078.

51. Berlin I, Grangé G, Jacob N, Tanguy ML. Nicotine patches in pregnant smokers: randomised, placebo controlled, multicentre trial of efficacy. BMJ. 2014;348:g1622.

52. Coleman T, Cooper S, Thornton JG, Grainge MJ, Watts K, Britton J, et al. A randomized trial of nicotine-replacement therapy patches in pregnancy. New England Journal of Medicine. 2012;366(9):808-18.

53. Oncken C, Dornelas E, Greene J, Sankey H, Glasmann A, Feinn R, et al. Nicotine gum for pregnant smokers: a randomized controlled trial. Obstet Gynecol. 2008;112(4):859-67.

54. Kapur B, Hackman R, Selby P, Klein J, Koren G. Randomized, double-blind, placebo-controlled trial of nicotine replacement therapy in pregnancy. Current Therapeutic Research, Clinical & Experimental. 2001;62(4):274-8.

55. Wisborg K, Henriksen TB, Jespersen LB, Secher NJ. Nicotine patches for pregnant smokers: a randomized controlled study. Obstet Gynecol. 2000;96(6):967-71.

56. El-Mohandes AAE, Windsor R, Tan S, Perry DC, Gantz MG, Kiely M. A randomized clinical trial of trans-dermal nicotine replacement in pregnant African-American smokers. Maternal and child health journal. 2013;17(5):897-906.

57. Hotham ED, Gilbert AL, Atkinson ER. A randomised-controlled pilot study using nicotine patches with pregnant women. Addict Behav. 2006;31(4):641-8.

58. Pollak KI, Oncken CA, Lipkus IM, Lyna P, Swamy GK, Pletsch PK, et al. Nicotine replacement and behavioral therapy for smoking cessation in pregnancy. American Journal of Preventive Medicine. 2007;33(4):297-305.

59. Coleman T, Chamberlain C, Davey M-A, Cooper Sue E, Leonardi-Bee J. Pharmacological interventions for promoting smoking cessation during pregnancy. Cochrane Database of Systematic Reviews. 2015(12):CD010078.

60. Vaz LR, Leonardi-Bee J, Aveyard P, Cooper S, Grainge M, Coleman T. Factors associated with smoking cessation in early and late pregnancy in the smoking, nicotine, and pregnancy trial: a trial of nicotine replacement therapy. nicotine & tobacco research. 2013;16(4):381-9.

61. Hollands GJ, Sutton S, McDermott MS, Marteau TM, Aveyard P. Adherence to and consumption of nicotine replacement therapy and the relationship with abstinence within a smoking cessation trial in primary care. nicotine & tobacco research. 2013;15(9):1537-44.

62. Matsui D. Adherence with drug therapy in pregnancy. Obstetrics and gynecology international. 2011;2012.

63. Bowker K, Campbell KA, Coleman T, Lewis S, Naughton F, Cooper S. Understanding pregnant smokers' adherence to nicotine replacement therapy during a quit attempt: a qualitative study. Nicotine & Tobacco Research. 2015;18(5):906-12.

64. Bowker K, Lewis S, Coleman T, Cooper S. Changes in the rate of nicotine metabolism across pregnancy: a longitudinal study. Addiction. 2015;110(11):1827-32.

65. Vine MF, Hulka BS, Margolin BH, Truong YK, Hu P-c, Schramm MM, et al. Cotinine concentrations in semen, urine, and blood of smokers and nonsmokers. American Journal of Public Health. 1993;83(9):1335-8.

66. Benowitz NL, Jacob P. Effects of cigarette smoking and carbon monoxide on nicotine and cotinine metabolism. Clinical Pharmacology & Therapeutics. 2000;67(6):653-9.

67. Kandel DB, Hu M-C, Schaffran C, Udry JR, Benowitz NL. Urine nicotine metabolites and smoking behavior in a multiracial/multiethnic national sample of young adults. American journal of epidemiology. 2007;165(8):901-10.

68. Rubinstein ML, Benowitz NL, Auerback GM, Moscicki A-B. Rate of nicotine metabolism and withdrawal symptoms in adolescent light smokers. Pediatrics. 2008;122(3):e643-e7.

69. Dempsey D, Jacob P, Benowitz NL. Accelerated metabolism of nicotine and cotinine in pregnant smokers. Journal of Pharmacology and Experimental Therapeutics. 2002;301(2):594-8.

70. Bullock H, Mountford J, Stanley R. Better Policy-Making London: Centre for Management and Policy Studies. 2001.

71. Björk B-C, Roos A, Lauri M, editors. Global annual volume of peer reviewed scholarly articles and the share available via different Open Access options. ELPUB2008; 2008.

72. Larsen PO, Von Ins M. The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. Scientometrics. 2010;84(3):575-603.

73. Wilson P, Petticrew M. Why promote the findings of single research studies? BMJ. 2008;336(7646):722-.

74. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1. 0: The Cochrane Collaboration; 2011 [updated March 2011]. Available from: www cochrane-handbook org. 2011.

75. Akobeng AK. Understanding systematic reviews and meta-analysis. Archives of disease in childhood. 2005;90(8):845-8.

76. Pai M, McCulloch M, Gorman JD, Pai N, Enanoria W, Kennedy G, et al. Systematic reviews and meta-analyses: an illustrated, step-by-step guide. The National medical journal of India. 2003;17(2):86-95.

77. Greenhalgh T, Thorne S, Malterud K. Time to challenge the spurious hierarchy of systematic over narrative reviews? European journal of clinical investigation. 2018;48(6).

78. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. Annals of internal medicine. 1997;126(5):376-80.

79. Mulrow CD. Rationale for systematic reviews. BMJ: British Medical Journal. 1994;309(6954):597.

80. Petticrew M. Why certain systematic reviews reach uncertain conclusions. British Medical Journal. 2003;326(7392):756.

81. Greenhalgh T. How to read a paper: getting your bearings (deciding what the paper is about). BMJ. 1997;315(7102):243-6.

82. Glass GV. Primary, secondary, and meta-analysis of research. Educational researcher. 1976;5(10):3-8.

83. Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. Systematic reviews in health care: meta-analysis in context. 2001;2:285-312.

84. Garattini S, Jakobsen JC, Wetterslev J, Bertelé V, Banzi R, Rath A, et al. Evidence-based clinical practice: Overview of threats to the validity of evidence and how to minimise them. European Journal of Internal Medicine. 2016;32:13-21.

85. Wetterslev J, Engstrøm J, Gluud C, Thorlund K. Response to "Comment by Higgins." Cochrane methods. Cochrane Database Syst Rev. 2012;2(suppl 1):33-5.

86. Imberger G, Thorlund K, Gluud C, Wetterslev J. False-positive findings in Cochrane meta-analyses with and without application of trial sequential analysis: an empirical review. BMJ Open. 2016;6(8):e011890.

87. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Fixed-Effect Versus Random-Effects Models. Introduction to Meta-Analysis: John Wiley & Sons, Ltd; 2009. p. 77-86.

88. Guo FR. Smoking links to the severity of Covid-19: An update of a meta-analysis. Journal of Medical Virology. 2020.

89. Higgins J, Thompson SG. Quantifying heterogeneity in a meta-analysis. Statistics in medicine. 2002;21(11):1539-58.

90. Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. New England Journal of Medicine. 1992;327(4):248-54.

91. Pogue JM, Yusuf S. Cumulating evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis. Controlled clinical trials. 1997;18(6):580-93.

92. Easterbrook PJ, Gopalan R, Berlin JA, Matthews DR. Publication bias in clinical research. The Lancet. 1991;337(8746):867-72.

93. Thorlund K, Engstrøm J, Wetterslev J, Brok J, Imberger G, Gluud C. User manual for trial sequential analysis (TSA). Copenhagen Trial Unit, Centre for Clinical Intervention Research, Copenhagen, Denmark. 2011;1:1-115.

94. Pogue J, Yusuf S. Overcoming the limitations of current meta-analysis of randomised controlled trials. The Lancet. 1998;351(9095):47-52.

95. Thorlund K, Imberger G, Walsh M, Chu R, Gluud C, Wetterslev J, et al. The number of patients and events required to limit the risk of overestimation of intervention effects in meta-analysis—a simulation study. PLoS ONE. 2011;6(10):e25491.

96. Imberger G, Gluud C, Boylan J, Wetterslev J. Systematic reviews of anesthesiologic interventions reported as statistically significant: problems with power, precision, and type 1 error protection. Anesthesia & Analgesia. 2015;121(6):1611-22.

97. Harrison W, Angoulvant F, House S, Gajdos V, Ralston SL. Hypertonic Saline in Bronchiolitis and Type I Error: A Trial Sequential Analysis. Pediatrics. 2018;142(3):e20181144.

98. Simmonds M, Salanti G, McKenzie J, Elliott J, Agoritsas T, Hilton J, et al. Living systematic reviews: 3. Statistical methods for updating meta-analyses. J Clin Epidemiol. 2017;91:38-46.

99. Borm GF, Donders ART. Updating meta-analyses leads to larger type I errors than publication bias. J Clin Epidemiol. 2009;62(8):825-30.

100. Guyatt GH, Mills EJ, Elbourne D. In the era of systematic reviews, does the size of an individual trial still matter? PLoS Med. 2008;5(1):e4.

101. Brok J, Thorlund K, Wetterslev J, Gluud C. Apparently conclusive meta-analyses may be inconclusive—trial sequential analysis adjustment of random error risk due to repetitive testing of accumulating data in apparently conclusive neonatal meta-analyses. International Journal of Epidemiology. 2009;38(1):287-98.

102. Jennison C, Turnbull BW. Group sequential methods with applications to clinical trials: CRC Press; 1999.

103. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. PLoS Med. 2007;4(3):e78.

104. Wetterslev J, Jakobsen JC, Gluud C. Trial Sequential Analysis in systematic reviews with meta-analysis. BMC Medical Research Methodology. 2017;17(1):39.

105. Nguyen T-L, Collins GS, Lamy A, Devereaux PJ, Daurès J-P, Landais P, et al. Simple randomization did not protect against bias in smaller trials. J Clin Epidemiol. 2017;84:105-13.

106. Lan GKK, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika. 1983;70(3):659-63.

107. Gehr BT, Weiss C, Porzsolt F. The fading of reported effectiveness. A meta-analysis of randomised controlled trials. BMC medical research methodology. 2006;6(1):25.

108. Thorlund K, Devereaux PJ, Wetterslev J, Guyatt G, Ioannidis JPA, Thabane L, et al. Can trial sequential monitoring boundaries reduce spurious inferences from meta-analyses? International Journal of Epidemiology. 2009;38(1):276-86.

109. Trikalinos TA, Churchill R, Ferri M, Leucht S, Tuunainen A, Wahlbeck K, et al. Effect sizes in cumulative meta-analyses of mental health randomized trials evolved over time. J Clin Epidemiol. 2004;57(11):1124-30.

110. Ioannidis JPA, Lau J. Evolution of treatment effects over time: empirical insight from recursive cumulative metaanalyses. Proceedings of the National Academy of Sciences. 2001;98(3):831-6.

111. Jackson D, Turner R. Power analysis for random-effects meta-analysis. Research Synthesis Methods. 2017;8(3):290-302.

112. Chai-Adisaksopha C, Thorlund K, Iorio A. Interpreting trial sequential analysis. Transfusion. 2016;56(12):2918-22.

113. Higgins JPT. Comment on "Trial sequential analysis: Methods and software for cumulative metaanalyses." Cochrane methods. Cochrane Database Syst Rev. 2012;2(suppl 1):32-3.

114. Jakobsen JC, Wetterslev J, Winkel P, Lange T, Gluud C. Thresholds for statistical and clinical significance in systematic reviews with meta-analytic methods. BMC medical research methodology. 2014;14(1):120.

115. Sterne JAC. Teaching hypothesis tests–time for significant change? Statistics in medicine. 2002;21(7):985-94.

116. Roloff V, Higgins J, Sutton AJ. Planning future studies based on the conditional power of a meta-analysis. Statistics in medicine. 2013;32(1):11-24.

117. Mascha EJ. Alpha, beta, meta: guidelines for assessing power and type I error in meta-analyses. LWW; 2015.

118. Kulinskaya E, Wood J. Trial sequential methods for meta-analysis. Research Synthesis Methods. 2014;5(3):212-20.

119. Claire R, Gluud C, Berlin I, Coleman T, Leonardi-Bee J. Using Trial Sequential Analysis for estimating the sample sizes of further trials: example using smoking cessation intervention. 2020.

120. Arain M, Campbell MJ, Cooper CL, Lancaster GA. What is a pilot or feasibility study? A review of current practice and editorial policy. BMC Medical Research Methodology. 2010;10(1):67.

121. Wittes J, Brittain E. The role of internal pilot studies in increasing the efficiency of clinical trials. Statistics in Medicine. 1990;9(1-2):65-72.

122. Thabane L, Ma J, Chu R, Cheng J, Ismaila A, Rios LP, et al. A tutorial on pilot studies: the what, why and how. BMC Medical Research Methodology. 2010;10(1):1.

123. Brok J, Thorlund K, Gluud C, Wetterslev J. Trial sequential analysis reveals insufficient information size and potentially false positive results in many meta-analyses. J Clin Epidemiol. 2008;61(8):763-9.

124. Naughton F, Prevost AT, Gilbert H, Sutton S, Naughton F, Prevost AT, et al. Randomized controlled trial evaluation of a tailored leaflet and

SMS text message self-help intervention for pregnant smokers (MiQuit). Nicotine & Tobacco Research. 2012;14(5):569-77.

125. Naughton F, Cooper S, Foster K, Emery J, Leonardi-Bee J, Sutton S, et al. Large multi-centre pilot randomized controlled trial testing a low-cost, tailored, self-help smoking cessation text message intervention for pregnant smokers (MiQuit). Addiction. 2017;112(7):1238-49.

126. Smoking and the young. A report of a working party of the Royal College of Physicians. London: RCP; 1992.

127. Batstra L, Hadders-Algra M, Neeleman J. Effect of antenatal exposure to maternal smoking on behavioural problems and academic achievement in childhood: Prospective evidence from a Dutch birth cohort. Early Hum Dev. 2003;75(1-2):21-33.

128. McAndrew F, Thompson J, Fellows L, Large A, Speed M, Renfrew MJ. Infant Feeding Survey 2010: Health and Social Care Information Centre: Health and Social Care Information Centre; 2012 [Available from: http://www.hscic.gov.uk/catalogue/PUB08694/Infant-Feeding-Survey-2010-Consolidated-Report.pdf. (Accessed: 14-12-2018.)]

129. Abroms LC, Ahuja M, Kodl Y, Thaweethai L, Sims J, Winickoff JP, et al. Text2Quit: results from a pilot test of a personalized, interactive mobile health smoking cessation program. Journal of Health Communication. 2012;17(sup1):44-53.

130. Abroms LC, Boal AL, Simmens SJ, Mendel JA, Windsor RA. A randomized trial of Text2Quit: a text messaging program for smoking cessation. American Journal of Preventive Medicine. 2014;47(3):242-50.

131. Free C, Whittaker R, Knight R, Abramsky T, Rodgers A, Roberts IG. Txt2stop: a pilot randomised controlled trial of mobile phone-based smoking cessation support. Tobacco Control. 2009;18(2):88-91.

132. Wetterslev J, Thorlund K, Brok J, Gluud C. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. J Clin Epidemiol. 2008;61(1):64-75.

133. Higgins JPT. Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. International Journal of Epidemiology. 2008;37(5):1158-60.

134. Thorlund K, Imberger G, Johnston BC, Walsh M, Awad T, Thabane L, et al. Evolution of heterogeneity (I2) estimates and their 95% confidence intervals in large meta-analyses. PLoS ONE. 2012;7(7):e39471.

135. Bauer P, Bretz F, Dragalin V, König F, Wassmer G. Twenty-five years of confirmatory adaptive designs: opportunities and pitfalls. Statistics in Medicine. 2016;35(3):325-47.

136. Committee CS. Should Cochrane apply error-adjustment methods when conducting repeated meta-analyses? ; 2018.

137. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. *Cochrane Handbook for Systematic Reviews of Interventions*: Cochrane; 2019. Available from: www.training.cochrane.org/handbook.

138. National Institute for Health and Clinical Excellence. Public health guidance 26: How to stop smoking in pregnancy and following childbirth. London; 2010 June 2010. Report No.: PH26.

139. Murin S, Rafii R, Bilello K. Smoking and smoking cessation in pregnancy. Clinics in chest medicine. 2011;32(1):75-91.

140. Oncken CA, Kranzler HR. What do we know about the role of pharmacotherapy for smoking cessation before or during pregnancy? Nicotine & Tobacco Research. 2009;11(11):1265-73.

141. Osadchy A, Kazmin A, Koren G. MOTHERISK ROUNDS: Nicotine Replacement Therapy During Pregnancy: Recommended or Not Recommended? Journal of Obstetrics and Gynaecology Canada. 2009;31(8):744-7.

142. Bruin JE, Gerstein HC, Holloway AC. Long-term consequences of fetal and neonatal nicotine exposure: a critical review. Toxicological sciences. 2010;116(2):364-74.

143. Benowitz NL, Dempsey DA, Goldenberg RL, Hughes JR, Dolan-Mullen P, Ogburn PL, et al. The use of pharmacotherapies for smoking cessation during pregnancy. Tobacco Control. 2000;9 Suppl 3:III91-III4.

144. Coleman T, Chamberlain C, Davey MA, Cooper SE, Leonardi-Bee J. Pharmacological interventions for promoting smoking cessation during pregnancy. Cochrane Database of Systematic Reviews. 2015(12).

145. Coleman T, Chamberlain C, Cooper S, Leonardi-Bee J. Efficacy and safety of nicotine replacement therapy for smoking cessation in pregnancy: systematic review and meta-analysis. Addiction. 2011;106(1):52-61.

146. West R, Hajek P, Stead L, Stapleton J. Outcome criteria in smoking cessation trials: proposal for a common standard. Addiction. 2005;100(3):299-303.

147. Lumley J, Chamberlain C, Dowswell T, Oliver S, Oakley L, Watson L. Interventions for promoting smoking cessation during pregnancy. The Cochrane database of systematic reviews. 2009(3):CD001055.

148. The Cochrane Collaboration. Review Manager (RevMan) [Computer program]. Version 5.3.,. Copenhagen: The Nordic Cochrane Centre 2014.

149. Schünemann H, Brozek J, Guyatt G, Oxman A. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. (Updated October 2013): McMaster University Toronto; 2013. [Available from: gdt.guidelinedevelopment.org/app/handbook/handbook.html. (Accessed: 14-08-2019.)]

150. Gradepro GDT. GRADEpro guideline development tool [software]. McMaster University. 2015.

151. Oncken C, Dornelas EA, Kuo C-L, Sankey HZ, Kranzler HR, Mead EL, et al. Randomized trial of nicotine inhaler for pregnant smokers. American journal of obstetrics & gynecology MFM. 2019;1(1):10-8.

152. Gould GS, Bovill M, Pollock L, Bonevski B, Gruppetta M, Atkins L, et al. Feasibility and acceptability of Indigenous Counselling and Nicotine (ICAN) QUIT in Pregnancy multicomponent implementation intervention and study design for Australian Indigenous pregnant women: a pilot cluster randomised step-wedge trial. Addict Behav. 2019;90:176-90.

153. ACTRN12618000972224. SISTAQUIT (Supporting Indigenous Smokers to Assist Quitting): a cluster randomised controlled trial to improve strategies for the management of smoking cessation in pregnant Aboriginal and/or Torres Strait Islander women 2018 [Available from: www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=374707. (Accessed: 01-08-2019.)]

154. IRCT20181209041904N1. Comparison of the effectiveness of treatment with nicotine patches versus support behavior in smoking cessation in pregnant women. 2018 [Available from: en.irct.ir/trial/35928. (Accessed: 01-08-2019.)]

155. Stout RL, Wirtz PW, Carbonari JP, Del Boca FK. Ensuring balanced distribution of prognostic factors in treatment outcome research. Journal of Studies on Alcohol, supplement. 1994(12):70-5.

156. Cooper S, Taggar J, Lewis S, Marlow N, Dickinson A, Whitemore R, et al. Effect of nicotine patches in pregnancy on infant and maternal outcomes at 2 years: follow-up from the randomised, double-blind, placebo-controlled SNAP trial. The Lancet Respiratory Medicine. 2014;2(9):728-37.

157. Squires J TE, Bricker D, Potter L. ASQ-3 User's Guide for the Ages and Stages Questionnaires. 3rd Edition ed. Baltimore: Paul H Brookes Publishing Co; 2009.

158. Verification SSoB. Biochemical verification of tobacco use and cessation. Nicotine & Tobacco Research. 2002;4(2):149-59.

159. Hartmann-Boyce J, Chepkin SC, Ye W, Bullen C, Lancaster T. Nicotine replacement therapy versus control for smoking cessation. Cochrane database of systematic reviews. 2018(5).

160. Dempsey DA, Benowitz NL. Risks and benefits of nicotine to aid smoking cessation in pregnancy. Drug safety. 2001;24(4):277-322.

161. Hickson C, Lewis S, Campbell KA, Cooper S, Berlin I, Claire R, et al. Comparison of nicotine exposure during pregnancy when smoking and abstinent with nicotine replacement therapy: systematic review and meta-analysis. Addiction. 2019;114(3):406-24.

162. Claire R, Coleman T, Leonardi-Bee J, Berlin I. Saliva cotinine concentrations in pregnant women who smoke and use nicotine patches. Addiction. 2019.

163. Kumar R, Gould GS. Tobacco Harm Reduction for Women Who Cannot Stop Smoking During Pregnancy—A Viable Option? JAMA pediatrics. 2019.

164. Lindson N, Chepkin SC, Ye W, Fanshawe TR, Bullen C, Hartmann-Boyce J. Different doses, durations and modes of delivery of nicotine replacement therapy for smoking cessation. Cochrane Database of Systematic Reviews. 2019(4).

165. Pereira TV, Ioannidis JPA. Statistically significant meta-analyses of clinical trials have modest credibility and inflated effects. J Clin Epidemiol. 2011;64(10):1060-9.

166. Wetterslev J, Thorlund K, Brok J, Gluud C. Estimating required information size by quantifying diversity in random-effects model meta-analyses. BMC medical research methodology. 2009;9(1):86.

167. Centers for Disease Control and P. Prevalence of Smoking during the last 3 months of Pregnancy, PRAMS, 2000–2008. 2013.

168. Blondel B, Kermarrec M. Les naissances en 2010 et leur évolution depuis 2003. Enquête nationale périnatale 2010. Paris, Inserm Éditions L'annonce prénatale d'un risque de handicap. 2011.

169. European Network for Smoking and Tobacco P. European Smoking Cessation Guidelines: The Authoritative Guide to a Comprehensive Understanding of the Implications and Implementation of Treatments and Strategies to Treat Tobacco Dependence. European Network for Smoking and Tobacco Prevention aisbl (ENSP) Belgium; 2011.

170. Dempsey D, Tutka P, Jacob P, Allen F, Schoedel K, Tyndale RF, et al. Nicotine metabolite ratio as an index of cytochrome P450 2A6 metabolic activity. Clinical Pharmacology & Therapeutics. 2004;76(1):64-72.

171. Ferguson SG, Gitchell JG, Shiffman S. Continuing to wear nicotine patches after smoking lapses promotes recovery of abstinence. Addiction. 2012;107(7):1349-53.

172. Berlin I, Singleton EG, Heishman SJ. A Comparison of the Fagerström Test for Cigarette Dependence and Cigarette Dependence Scale in a Treatment-seeking sample of pregnant smokers. Nicotine & Tobacco Research. 2015;18(4):477-83.

173. Pickett KE, Wakschlag LS, Dai L, Leventhal BL. Fluctuations of maternal smoking during pregnancy. Obstet Gynecol. 2003;101(1):140-7.

174. Pickett KE, Rathouz PJ, Kasza K, Wakschlag LS, Wright R. Self-reported smoking, cotinine levels, and patterns of smoking in pregnancy. Paediatr Perinat Epidemiol. 2005;19(5):368-76.

175. Dienes Z. Make the most of your data with Bayes [Available from: http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/Bayes. (Accessed: 08-05-2019.)]

176. Shahab L, Beard E, Brown J, West R. Prevalence of NRT use and associated nicotine intake in smokers, recent ex-smokers and longer-term ex-smokers. PLoS ONE. 2014;9(11):e113045.

177. Benowitz NL, Jacob Iii P, Ahijevych K, Jarvis MJ, Hall S, LeHouezec J, et al. Biochemical verification of tobacco use and cessation. Nicotine & Tobacco Research. 2002;4(2).

178. Claire R, Chamberlain C, Davey MA, Cooper SE, Berlin I, Leonardi-Bee J, et al. Pharmacological interventions for promoting smoking cessation during pregnancy. Cochrane Database of Systematic Reviews. 2020(3).

179. Mahmoud AN, Gad MM, Elgendy AY, Elgendy IY, Bavry AA. Efficacy and safety of aspirin for primary prevention of cardiovascular events: a meta-analysis and trial sequential analysis of randomized controlled trials. Eur Heart J. 2019;40(7):607-17.

180. Bolland MJ, Grey A, Avenell A. Effects of vitamin D supplementation on musculoskeletal health: a systematic review, meta-analysis, and trial sequential analysis. The Lancet Diabetes & Endocrinology. 2018;6(11):847-58.

181. National Institute for Health and Clinical Excellence. Quitting smoking in pregnancy and following childbirth. NICE Public Health Guidance 26. London; 2010. Report No.: PH26.

182. Makrgeorgou A, Leonardi-Bee J, Bath-Hextall FJ, Murrell DF, Tang MLK, Roberts A, et al. Probiotics for treating eczema. Cochrane Database of Systematic Reviews. 2018(11).

183. Uthman OA, Nduka CU, Abba M, Enriquez R, Nordenstedt H, Nalugoda F, et al. Comparison of mHealth and face-to-face interventions for smoking cessation among people living with HIV: Meta-analysis. JMIR mHealth and uHealth. 2019;7(1):e203.

184. Wang J-H, van Haselen R, Wang M, Yang G-L, Zhang Z, Friedrich ME, et al. Acupuncture for smoking cessation: A systematic review and meta-analysis of 24 randomized controlled trials. Tobacco induced diseases. 2019;17.

Appendix A: Tables and figures from systematic review (Chapter 5)

Systematic review: characteristics of included studies tables

Berlin 2014

Methods	Double-blind, placebo-controlled, parallel-group RCT
Participants	476 pregnant women aged ≥ 18 years, between 9 and 20 weeks' gestation who smoked at least 5 daily cigarettes and scored at least 5 on a scale measuring motivation for quitting smoking (range 0 to 10)
Interventions	Intervention and control differed only in the provision of active or visually identical placebo transdermal patches. The intervention patch delivered nicotine as nicotine replacement therapy over a 16-hour period. Both 10 mg and 15 mg patches were used, and women's doses ranged from 10 mg to 30 mg per day. A saliva sample was collected at the woman's first trial visit/contact with the research team. Between this and a second visit/contact, which occurred 2 weeks later, women were instructed to either stop smoking or to reduce this to less than 5 daily cigarettes. Women who managed to reduce or stop smoking in this way were, at their second visit, randomised to either placebo or active patch in a 1:1 ratio. The nicotine dose used for women's first prescription of NRT (made at this 2nd trial visit) was based on their saliva cotinine level obtained from the sample given at visit 1 with the aim being to attempt 100% substitution of nicotine obtained from smoking for that obtained via patches. Women were instructed to use NRT from their quit date until delivery. Smoking and using patches was not encouraged (this is described as a "safety concern"). However, if women did have a temporary lapse to smoking, they were allowed to remain on NRT afterwards. Both groups received counselling on how to use patches.
Outcomes	There were 2 primary outcomes, 1 maternal and 1 relating to infants: complete, continuous abstinence from smoking since the quit date and infant birthweight. A positive abstinence outcome was recorded where women self-reported 7 days abstinence from smoking at each study visit, and this was confirmed by an exhaled CO reading of 8 ppm or less. There were up to 7 study visits with the final visit intended for 1 month prior to delivery; no lapses to smoking were permitted.

Notes	The cessation outcome used was more stringent than in many studies; often some allowance for temporary lapses to smoking is permitted, and many studies assess smoking status as a smaller number of time points in pregnancy. Dates of study: October 2007 to January 2013 Funding sources: "This study was funded by the Ministry of Health, France (grant No MA05 00150) and co-sponsored by Assistance publique-Hôpitaux de Paris (P060604).The Ministry of Health and Assistance publique-Hôpitaux de Paris had no role in the design and conduct of the study; the collection, conduct, analysis, and interpretation of the data; or the preparation, review, or approval of the manuscript." Gunnar Gustavsson and McNeil-Johnson & Johnson provided the nicotine and placebo patches free of charge. Declarations of interest: "All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that: none had support of any kind for the submitted work; IB has served as a paid consultant for Pfizer, Novartis, and Ethypharm in the past three years; none of the authors' spouses, partners, or children has financial relationships that may be relevant to the submitted work; and none of the authors has non-financial interests that may be relevant to the submitted work."

Coleman 2012

Methods	Double-blind, placebo-controlled RCT – stratified by trial centre only
Participants	Pregnant women (n = 1050) who agreed to set a quit date, were 16 to 50 years of age, were at 12 to 24 weeks of gestation, smoked 10 or more cigarettes daily before pregnancy, currently smoked 5 or more cigarettes daily, and had an exhaled CO concentration of at least 8 ppm
Interventions	Intervention and control conditions differed only in the provision of transdermal patches; the intervention group received active patches and the control group received placebo patches. Research midwives were trained to provide behavioural support according to national standards, with the use of a manual that included guidance from a British expert trainer of smoking-cessation professionals and behavioural approaches from the Smoking Cessation or Reduction in Pregnancy Treatment trials that were believed to be relevant to British people who smoke. At enrolment, research midwives provided behavioural support lasting up to 1 h, and participants agreed to a quit date within the following 2 weeks; follow-up was timed from the quit date. Subsequently, participants were randomly assigned to receive a 4-week supply of transdermal patches for NRT (at a dose of 15 mg per 16 h) or visually identical placebos, which were started on the quit date (all study treatment was purchased at market rates from United Pharmaceuticals). 1 month after the quit date, women who were not smoking, as validated by an exhaled CO concentration of less than 8 ppm, were issued another 4-week supply of patches. In addition to behavioural support at enrolment, research midwives provided 3 sessions of behavioural support by telephone to participants: 1 session on the quit date, 1 session 3 days afterward, and 1 session at 4 weeks. The women who collected a 2nd month's supply of nicotine-replacement or placebo patches also received face-to-face support from the research midwife at the time of collection. Women were offered additional support from local National Health Service smoking cessation services and were encouraged to ask for support from the research midwives or smoking cessation service staff; support was provided according to the manual.

Outcomes	Prolonged smoking cessation between a quit date soon after enrolment and delivery, validated by both exhaled CO monitoring and saliva cotinine estimation. Cut points: exhaled CO, smoking was defined as > 7 ppm; saliva cotinine, smoking defined as > 9 ng/dL. Birth outcomes including Apgar score at 5 min after birth, cord arterial blood pH, intraventricular haemorrhage, neonatal convulsions, congenital abnormalities, necrotising enterocolitis, mechanical ventilation of infant, assisted vaginal delivery, maternal death, and caesarean section. For infants: survival to 2 years of age without developmental impairment, reported respiratory symptoms. Maternal: self-reported abstinence from smoking for at least 7 days reported at 6, 12, and 24 months after childbirth, prolonged abstinence from smoking since a quit date set in pregnancy and until 24-month follow-up (defined as having validate abstinence at delivery followed by reported abstinence at all outcome points listed above).
Notes	Dates of study: May 2007 to February 2010 Funding sources: "Supported by a grant from the NIHR Health Technology Assessment Programme (06/07/01)" Declarations of interest: "No potential conflict of interest relevant to this article was reported."

El-Mohandes 2013

Methods	Non-placebo, parallel-design RCT
Participants	52 English-speaking pregnant women who smoked and were residents of Washington, DC in the USA, of ethnic minority backgrounds, aged at least 18 years, and less than 30 weeks' gestation. Women needed to express a desire to quit and have an expired-air CO reading of 8 ppm or less and a salivary cotinine of 20 ng/mL or less (NB: ClinicalTrials.gov website says 30 ng/mL or less) or a urinary cotinine of 100 ng/mL or less.
Interventions	1:1 ratio randomisation, stratified by site and initial salivary cotinine levels to either 1) cognitive behavioural therapy (CBT) and NRT transdermal patches or 2) CBT alone. NRT: a 10-week course of 24-hour patches was offered, with initial dosing varying with baseline salivary cotinine measurements. Women with levels of ≥ 100 ng/mL were issued 21 mg patches for 2 weeks, 14 mg patches for 4 weeks, and finally 7 mg patches for 4 weeks. Women with levels of ≥ 20 ng/mL and ≤ 100 ng/mL were issued 14 mg patches for 6 weeks and 7 mg patches for 4 weeks. The first batch of patches was issued at the 2nd study visit at which salivary cotinine levels were available. Participants were given clear verbal and written instructions on patch use. They were advised never to smoke whilst using the patch, to remove the patch before going to sleep, and not to use other NRT concurrently. CBT: this was the same for both groups.
Outcomes	Smoking cessation outcome: during the study participants made 6 visits to the study team in the antenatal period. At visit 2 (V2), trial interventions were initiated, and at each of visits V3 to V6 (the last before childbirth), women were asked if they had smoked since their previous clinic visit (e.g. at V3, they were asked if they had smoked since V2). Participants who reported smoking cessation had this validated using exhaled CO, with abstinence viewed as confirmed by a reading of < 8 ppm. The trial manuscript reports point prevalence of abstinence from smoking at each time point, and

	data from V6 are used in analyses. All data were validated (self-report not available), but the period of abstinence that was validated is unclear and varied with the interval between clinic visits. Secondary outcomes reported in the trial manuscript: premature birth (i.e. at < 37 weeks' gestation); gestational age at birth; mean birthweight and low birthweight < 2500 g. The following outcomes were also collected, as clarified by the authors: ability to not smoke for 24 h or more; longest number of days that the woman was able to go without even a puff of smoking; frequency of smoking at least puff during the last 7 days; number of cigarettes smoked each day; number of cigarettes smoked during the past 24 h; and frequency of use of other forms of tobacco.
Notes	Title of paper states that it was conducted in "African-American smokers", but in manuscript participants are described as "ethnic minority women", and inclusion criteria on ClinicalTrials.gov includes Hispanic women. Dates of study: July 2006 to May 2010 Funding sources: "This study was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (U10 HD036104 and U18 HD031206-07). This research was supported, in part, by the intramural program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development." Declarations of interest: "None of the authors have any competing interests to declare."

Hotham 2006

Methods	Non-placebo, parallel-design RCT
Participants	40 healthy Australian women between 12 and 28 weeks' pregnant and smoking >= 15 cigarettes daily with an exhaled breath CO reading of > 8 ppm
Interventions	Control group: 5-minute counselling at baseline and further brief counselling (< 2 minutes' duration) at follow-up visits. Intervention: counselling as above plus an element concerning correct use of NRT plus 15 mg/16-hour patches for a maximum of 12 weeks.
Outcomes	Smoking cessation (point prevalence) at final antenatal visit. Women seen "at least monthly during gestation"; also seen within 48 h of delivery when exhaled CO and saliva sample (for cotinine) taken and by telephone at 6 weeks and 3 months.
Notes	Exhaled CO readings used to validate point prevalence cessation at final antenatal visit. Cut point = 8 ppm CO. Author clarification used to obtain this information as not clear in research report. No data on smoking outcomes after childbirth are reported in the manuscript. Dates of study: not reported Funding sources: "This pilot study was supported by the Health Promotion Branch of the (then) South Australian Health Commission, now the Department of Health (SA). The WCH Perinatal Pathology Fund funded cotinine tests, performed using a competitive micro-plate immuno-assay (COTININE MICRO-PLATE EIA)." Declarations of interest: not reported

Kapur 2001

Methods	Parallel-design RCT with active and placebo patches and clinicians/researchers and participants unaware of allocation
Participants	30 healthy Canadian women between 12 and 24 weeks' pregnant and smoking >= 15 cigarettes daily who want to quit smoking and could not do so in 1st trimester
Interventions	12-week course of NRT or identical placebo patches: 15 mg/18-hour patch for 8 weeks, then 10 mg/18-hour patch for 2 weeks, and finally 5 mg/18-hour patch for 2 weeks. Behavioural counselling at baseline and at all follow-up points. Counselling at baseline included a video explaining how to use patch; also counselling at all follow-ups. Weekly telephone contact with women. Intervention = active patch, control = placebo
Outcomes	Smoking cessation (unclear if point prevalence or continuous cessation measured) 8 weeks into programme (20 to 32 weeks into pregnancy). Follow-up also at weeks 1 and 4 into programme with saliva and serum cotinine measured at all time points.
Notes	Primary outcome validated at 8 weeks into programme. Cotinine cut point not reported, but paper states that "in no case was smoking cessation associate with thiocyanate levels of > 1 ug/ml". Dates of study: not reported Funding sources: "This study was supported by a grant from the Canadian Institutes of Health Research (CIHR)." Declarations of interest: "Gideon Koren, MD, is a Senior Scientist of the CIHR."

Oncken 2008

Methods	Parallel-design RCT with active and placebo NRT gum and clinicians/researchers and participants unaware of allocation
Participants	194 healthy, US English-/Spanish-speaking women <= 26 weeks' pregnant, smoking >= 1 cigarette daily and aged >= 16 years
Interventions	12 weeks treatment with either 2 mg NRT gum or identical placebo. 6 weeks full treatment was followed by 6 weeks tapering of treatment. Instructed not to chew > 20 pieces daily and to use 1 piece of gum for each substituted cigarette. Additionally, all participants received individual counselling at baseline and at all 8 follow-ups: 2, 35-minute counselling sessions at baseline and within 1 week of quit date and shorter sessions at other follow-ups. Intervention = active gum, control = placebo
Outcomes	Self-reported 7-day point prevalence abstinence at 6 weeks after treatment commenced, at 32 to 35 weeks of pregnancy, and at 6 to 12 weeks after delivery. Exhaled CO of less than 8 ppm used for validation all time points.
Notes	Dates of study: July 2003 to April 2007 Funding sources: "Supported by NIH grants R01 DA15167, GCRC grant M01 RR006192, P50 DA013334, P50 AA015632. Nicotine Gum was provided free of charge from Glaxo-Smith Kline." Declarations of interest: "Dr. Oncken has received consulting fees and honoraria from Pfizer (New York, NY) for advisory board meetings. She has received at no cost nicotine and/or placebo products from Glaxo-SmithKline (Philadelphia, PA) for smoking cessation studies (i.e., for pregnant women, postmenopausal women). She has received grant funding from Pfizer for smoking cessation studies and from Nabi Biopharmaceuticals (Boca Raton, FL) for a nicotine vaccine study. Dr. Kranzler has received consulting fees from Ortho-McNeil Pharmaceuticals (Raritan, NJ), H. Lundbeck A/S (Copenhagen, Denmark), Forest Pharmaceuticals (St. Louis, MO), elbion NV (Leuven, Belgium),

	Sanofi-Aventis (Bridgewater, NJ), Solvay Pharmaceuticals (Bruxelles, Belgium), and Alkermes, Inc. (Cambridge, MA). He has received research support from Ortho-McNeil Pharmaceuticals and Bristol-Myers Squibb Company (New York, NY), and honoraria from Forest Pharmaceuticals and Alkermes, Inc. The other authors have no potential conflicts of interest to disclose."

Oncken 2019

Methods	Parallel-design RCT with active and placebo NRT inhaler and clinicians/researchers and participants unaware of allocation
Participants	137 healthy US English-/Spanish-speaking women smoking at least 5 cigarettes per day, 13 to 26 weeks' gestation, ≥ 16 years of age, intending to carry their pregnancy to term, and living in a stable residence
Interventions	6 weeks' treatment using NICOTROL inhaler (nicotine inhalation system) delivering 4 mg of nicotine from a porous plug containing 10 mg nicotine. Participants were encouraged to continue the use of the inhaler as long as they were actively trying to quit smoking. Participants instructed to puff on the inhaler 3 to 4 times per minute for up to 20 minutes and to inhale deeply in short breaths as they would normally smoke a cigarette. Participants who smoked ≥ 10 CPD were instructed to begin with 4 to 12 cartridge inhalers per day; women who smoked 5 to 9 CPD were instructed to begin with 1 to 4 cartridge inhalers per day, based on an estimated 1 to 2 mg of nicotine delivery per cigarette, with each cartridge inhaler estimated to release 4 mg of nicotine. At baseline and 1 week after quit date, participants received 35 minutes of individual smoking cessation counselling by a study nurse trained to deliver the counselling using a motivational interviewing approach. Intervention = nicotine inhaler, control = placebo
Outcomes	Self-reported 7-day point prevalence abstinence at 6 weeks after quit date, at 32 to 36 weeks of pregnancy, and at 1 and 6 months after delivery. Exhaled CO of less than 4 ppm used for validation at all time-points.
Notes	Study planned to recruit 360 women, but the trial was stopped after a recommendation from the Data and Safety Monitoring Board due to futility in detecting differences in the primary outcome. Dates of study: August 2012 to January 2017

	Funding sources: "This study was supported by National Institutes of Health (NIH) of United States grant R01HD069314 and the Lowell P. Weicker Clinical Research at the University of Connecticut School of Medicine. The study medication was donated by Pfizer Pharmaceuticals." Declarations of interest: "Dr Kranzler is a member of the American Society of Clinical Psychopharmacology's Alcohol Clinical Trials Initiative, which was supported in the last 3 years by AbbVie, Alkermes, Ethypharm, Indivior, Lilly, Lundbeck, Otsuka, Pfizer, Arbor, and Amygdala Neurosciences and is named as an inventor on Patent Cooperation Treaty patent application 15/878,640 entitled genotype-guided dosing of opioid agonists, filed Jan. 24, 2018. The other authors report no conflict of interest."

Pollak 2007

Methods	Non-placebo, parallel-design RCT
Participants	181 healthy US English-speaking women between 13 and 25 weeks' pregnant, smoking >= 5 cigarettes daily, and aged >= 18 years. Must have smoked > 100 cigarettes in lifetime.
Interventions	Control group: 5 face-to-face and 1 telephone behavioural counselling sessions with booklet and support materials. Intervention group: counselling as above but with additional focus on use of NRT. Women permitted choice of NRT from patch, gum, or lozenge. Patch dose depended on CPD: < 10 CPD, 7 mg/16 h; 10 to 14 CPD, 14 mg/16 h; >= 15 CPD, 21 mg/16 h. Where gum or lozenge was used, one 2 mg piece was used for each cigarette smoked daily. Maximum of 6 weeks' NRT provided, and no NRT provided when women returned to smoking.
Outcomes	Self-reported 7-day point prevalence abstinence at 38 weeks. Also follow-up at 7 weeks after randomisation and 3 months' postpartum using self-report data. Saliva samples for cotinine validation were collected at the intervention session that coincided with each telephone survey from all women regardless of smoking status. Cut point for primary outcome <= 10 ng/mL. Validation data were collected at all 3 time points, but are only reported for the 2 data collection points within pregnancy.
Notes	Choices of NRT: 72/122 patch = 59%, 32/122 gum = 26.2% and 12/122 lozenge = 9.8%. 19 women chose another formulation as they could not quit with initial selection (changes not recorded). Dates of study: May 2003 to August 2005 Funding sources: "This work was supported by the National Cancer Institute (grant R01CA089053 and operated under IND #67,259)." NRT donated by GlaxoSmithKline. Declarations of interest: "No financial disclosures were reported by the authors of this paper."

Wisborg 2000

Methods	Parallel-design RCT with active and placebo patches and clinicians/researchers and participants unaware of allocation
Participants	250 healthy Danish women < 22 weeks' pregnant and smoking >= 10 cigarettes daily
Interventions	11-week course of NRT or identical placebo patches: 15 mg/16 h for 8 weeks then 10 mg/16 h for 3 weeks plus behavioural counselling and information pamphlet. Intervention = active patch, control = placebo
Outcomes	Self-reported abstinence of >= 7 days at 2nd, 3rd, and 4th prenatal visits (4 weeks prior to delivery). Follow-ups at times above and also by telephone at 3 months and 1 year after delivery.
Notes	Saliva cotinine level < 26 ng/mL at the 4th visit (4 weeks prior to expected delivery date) used to validate reported smoking cessation. The test used could not detect lower than 20 ng/mL (data verified by communication with author). Only self-report data were collected after childbirth. Dates of study: October 1995 to October 1997 Funding sources: "This study was supported by the Danish Cancer Society and the Ministry of Health (The National Health Fund supported this study for Research and Development). Pharmacia & Upjohn provided nicotine patches." Declarations of interest: not reported

Systematic review: Table of adherence with NRT regimens

Study	Adherence with offered regimen as a percentage of complete course	Adherence with offered regimen in terms of period of use
Wisborg 2000	Complete adherence with 11-week course: nicotine group = 11%, placebo = 7%. Partial adherence (up to 8 weeks' use): nicotine group = 17%, placebo = 8%.	Median number patches (ranges): nicotine group = 14 (0 to 77), median = approximately 2 weeks; placebo = 7 (0 to 77), median = approximately 1 week.
Kapur 2001	In the nicotine group, 4/17 (23.5%) completed the 14-week programme. In the placebo group, no participants completed the programme.	In the nicotine group, 4/17 (23.5%) completed the 14-week programme; 3/17 (17.6%) used the patch for at least 3 weeks; and 10/17 (58.8%) used the patch for less than 1 week. In the placebo group, no participants completed the programme; 3/13 (23%) used the patch for between 4 and 5 weeks; and 10/13 (76.9%) used the patch for < 1 week.
Hotham 2006	25% (5) participants complied fully with protocol: "continuous patch use till 12 weeks or confident that abstinence achieved or adverse reaction experienced".	50% (10) of participants used NRT for 6 weeks or less.
Pollak 2007	Difficult to ascertain from manuscript. A secondary publication reported that 29% of participants used NRT as directed for intended 6-week programme.	Means of reported periods of use: Patch = 23.4 patches = 3.3 weeks Gum = 8 days Lozenge = 4 days
Oncken 2008	Not clearly reported.	The nicotine group used gum for a mean (SD) of 37.8 (3.8) days (i.e. just > 5 weeks). The placebo group used gum for a mean (SD) of 29.9 (3.4) days (i.e. just > 4 weeks).

Coleman 2012	Limited compliance with the intervention. Only 7.2% of women (35 of 485) assigned to receive NRT and 2.8% (14 of 496) assigned to receive placebo reported using trial medications for more than 1 month (2 months represented a complete course); rates of use of non-study NRT were very low. Most participants had no additional contact, either face-to-face or by text message, with smoking cessation advisors; amongst those who did, the frequency of contact was similar in the 2 groups.	Most participants discontinued patches after using them for only a short period: in the nicotine group 60.1% of participants used patches for no longer than 2 weeks, whilst in the placebo patch group this figure was 76.8%.
Berlin 2014	In contrast to other studies, women were issued with a much longer course of transdermal patches, i.e. from women's quit dates to their delivery. Compliance was measured using self-reported data on patches used between study visits and was obtained at 1016 study visits from 307 (76%) participants: 164 (84%) in the NRT group and 143 (72%) in the placebo group. Median (IQR) reported patch use was 85% (56% to 99%) in the NRT group and 83% (56% to 95%) in the placebo group. However, it is not clear how these figures relate to the rate with which participants discontinued the intervention. Overall, 225 (60.0%) of participants stopped using trial treatments: 105 (51.7%) in the NRT group and 60.3% in the placebo group.	This was not reported, but it has less meaning for this RCT, as women started using patches at different points in pregnancy and continued until childbirth.
Oncken 2019	Not clearly reported.	The nicotine group used the inhaler for a mean (SD) of 36.39 (23.92) days (i.e. just > 5 weeks) and used a mean (SD) of 1.70 (1.19) cartridges per day. The placebo group used the inhaler for a mean (SD) of 34.11 (20.54) days (i.e. just < 5 weeks) and used a mean (SD) of 1.81 (1.62) cartridges per day.

		Neither of these were statistically significant differences between groups (number of days, P = 0.587; number of cartridges, P = 0.701). Compliance with the inhaler during treatment was 69% in the placebo group and 70% in the nicotine group.

Forest plots

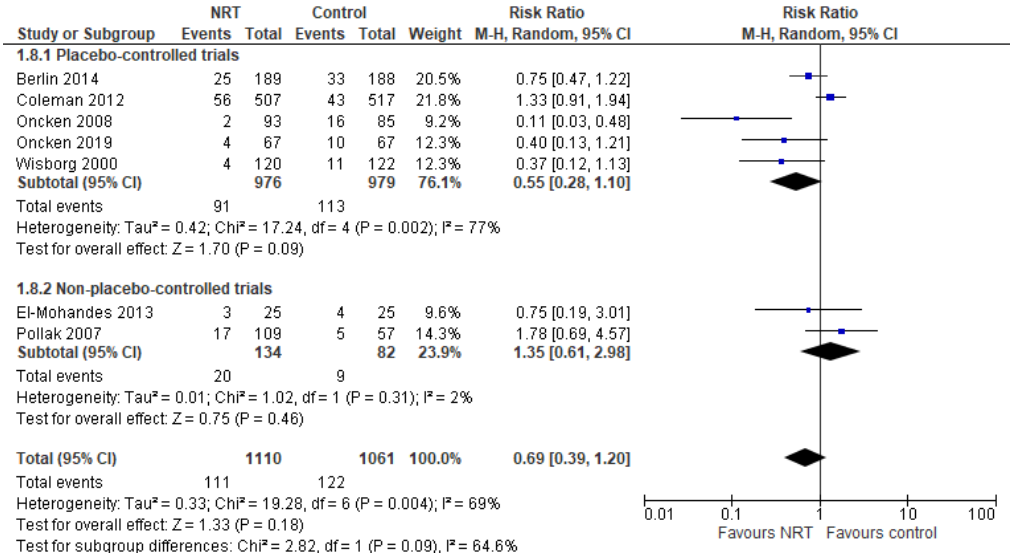

Figure S1 Forest plot of nicotine replacement therapy versus control, outcome: Low birthweight (< 2500 g)

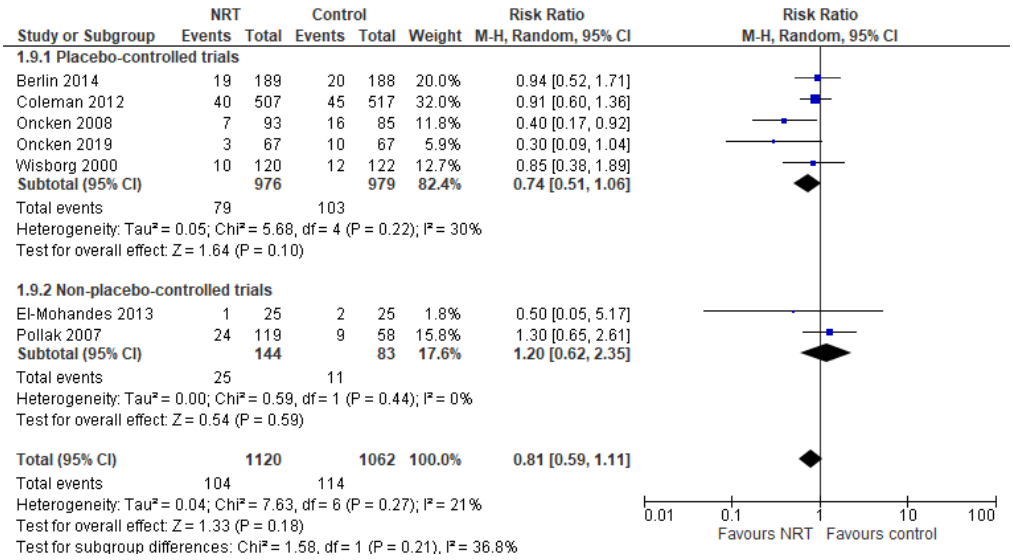

Figure S2 Forest plot of nicotine replacement therapy versus control, outcome: Preterm birth (birth < 37 weeks)

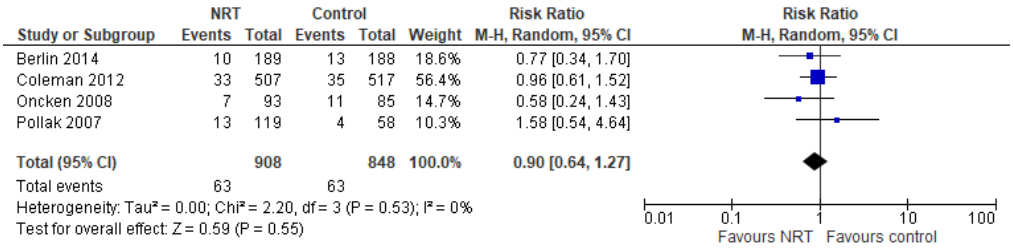

Figure S3 Forest plot of nicotine replacement therapy versus control, outcome: Neonatal intensive care unit admissions.

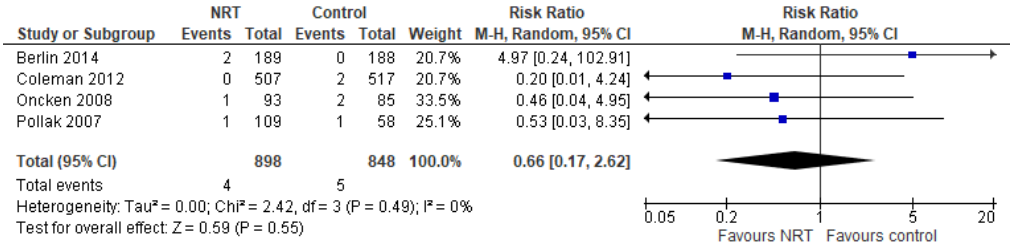

Figure S4 Forest plot of nicotine replacement therapy versus control, outcome: Neonatal death.

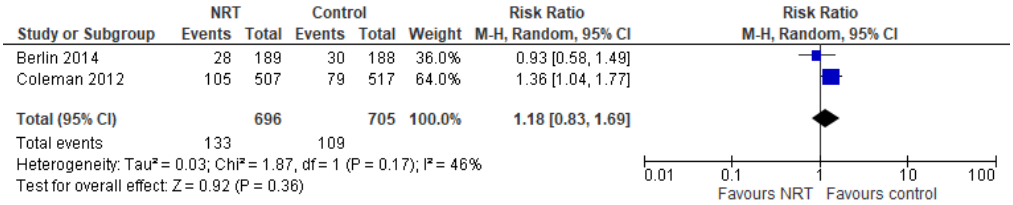

Figure S5 Forest plot of nicotine replacement therapy versus control, outcome: Caesarean section.

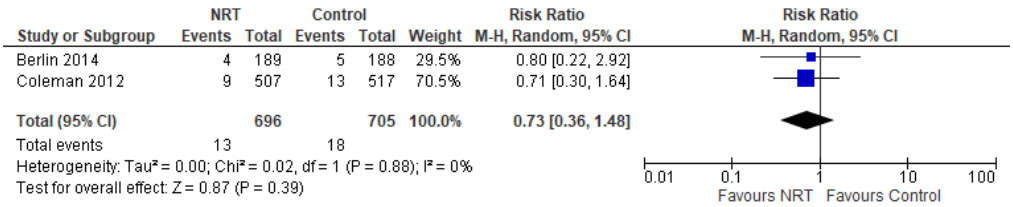

Figure S6 Forest plot of nicotine replacement therapy versus control, outcome: Congenital abnormalities.

Appendix B

Publications

Saliva cotinine concentrations in pregnant women who smoke and use nicotine patches

ADDICTION
SSA SOCIETY FOR THE STUDY OF ADDICTION
RESEARCH REPORT
doi:10.1111/add.14662

Saliva cotinine concentrations in pregnant women who smoke and use nicotine patches

Ravinder Claire[1], Tim Coleman[1], Jo Leonardi-Bee[2] & Ivan Berlin[3,4,5]

Division of Primary Care, University of Nottingham, Nottingham, UK,[1] Division of Epidemiology and Public Health, University of Nottingham, Nottingham, UK,[2] Département de pharmacologie, Hôpital Pitié-Salpêtrière, Paris, France,[3] Centre Universitaire de Médecine Générale et Santé publique, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland[4] and CESP-INSERM U1018, Paris, France[5]

ABSTRACT

Background and Aims Due to concerns about increased exposure to nicotine, pregnant women using nicotine replacement therapy (NRT) to stop smoking are usually advised to stop using NRT if they relapse to smoking. This study investigated whether this is justified. We compared changes in saliva cotinine from baseline to 2 weeks post-target quit date pregnant smokers who relapsed to smoking and continued to use their patches having been assigned to use nicotine patches or placebo. **Design and Setting** Controlled pre–post design stratified by intervention condition from the 'Study of Nicotine Patch in Pregnancy', a randomized, placebo-controlled trial. **Participants** A sample of 268 pregnant women, assigned placebo ($n = 122$) or nicotine ($n = 146$) patches, who returned for further supplies of patches and who reported any smoking in the week prior to a visit at 2 weeks after their target quit date. **Measurements** Saliva cotinine concentrations were measured at baseline and 2 weeks after participants' target quit dates. Any smoking in the previous week was assessed by self-report, validated by expired air carbon monoxide (CO). **Findings** There was no change in saliva cotinine concentrations between baseline and 2 weeks post-target quit date in saliva cotinine concentration in the nicotine patch group [ratio of geometric means = 0.94, 95% confidence interval (CI) = 0.83 to 1.07; $P = 0.37$, Bayes factor = 0.15]. However, there was a reduction in reported number of cigarettes smoked/day (mean difference −6, 95% CIs −7 to −5, $P < 0.001$) and in CO concentrations (mean difference −3.0 parts per million, 95% CIs −4.2 to −1.9, $P < 0.001$). These changes were not significantly different from changes in the placebo group except for cigarette consumption, which reduced more in the nicotine group ($P = 0.046$). **Conclusions** In women trying to stop smoking with the aid of a nicotine patch but having smoked at 2 weeks post-target quit, their nicotine concentration did not change from baseline, but they reported smoking fewer cigarettes and had lower carbon monoxide concentrations.

Keywords Cotinine, nicotine, nicotine replacement therapy, pregnancy, smoking, smoking cessation.

Correspondence to: Ravinder Claire, Division of Primary Care, University of Nottingham, Room 1502 Tower Building, Nottingham NG7 2RD, UK. E-mail: ravinder.claire@nottingham.ac.uk

Submitted 18 December 2018; initial review completed 1 February 2019; final version accepted 14 May 2019

INTRODUCTION

Smoking during pregnancy is the leading modifiable risk factor for poor maternal and infant health outcomes. Pregnancy-related health problems associated with smoking during pregnancy include complications during labour, increased risk of miscarriage, premature birth, stillbirth and low birth weight [1–3]. Despite this, approximately 12% of pregnant women in the United Kingdom, 13% in the United States and 20% in France continue to smoke during pregnancy [4–6]. Several national guidelines have adopted using nicotine replacement therapy (NRT) for supporting pregnant smokers to quit, based on the idea that NRT is probably safer than smoking as it does not contain the toxins present in tobacco smoke [7,8].

While NRT has been proven to be effective in non-pregnant smokers [9], its efficacy in pregnancy is uncertain [10]. The reason for this uncertainty is unclear; however, it is hypothesized that physiological changes in pregnancy could affect nicotine's metabolism [11]. Potential factors for the increased metabolism rate include a higher concentration or activity of metabolic enzymes involved and increased blood flow through the liver during pregnancy [12]. Cotinine is the principal metabolite of nicotine, and the clearance of nicotine and cotinine is 60 and

140% higher, respectively, during pregnancy [13]. An increase in metabolic rate could signify that nicotine supplied through standard dose NRT may be insufficient to alleviate smoking withdrawal symptoms in pregnancy and to provide therapeutic effects.

A systematic review and meta-analysis comparing nicotine exposure in pregnant women when smoking, and their nicotine exposure when abstinent and using NRT, found that NRT exposes women to lower doses of nicotine than does smoking [14]. Generally, in studies included in this review, such as the Smoking, Nicotine and Pregnancy (SNAP) trial, women were instructed to discontinue use of nicotine patches if they had even brief smoking lapses [15]. This mimics routine health care, where pregnant women are usually advised to stop using NRT if they lapse to smoking, even for short periods. There is concern that concomitant smoking and NRT use could increase exposure to nicotine and potentially more tobacco smoke toxins if they smoked heavily when using NRT. However, in pregnancy this assumption is untested, and we know little about women's smoking behaviour when they use NRT concurrently. This is important, as women who lapse to smoking may still want to quit. In a non-pregnant population, continued use of nicotine patches has been found to promote recovery from lapses [16]; if this is the case during pregnancy, women may have better chances of cessation if NRT is continued.

This study aims to investigate and compare: (1) changes in saliva cotinine and other indicators of smoking intensity in women using nicotine or placebo patches and smoking concurrently with those when they only smoked; and (2) whether these changes differed between nicotine and placebo patch use.

METHODS

Design

This is a secondary analysis of data from the Study of Nicotine Patch in Pregnancy (SNIPP) [17]. SNIPP was a multi-centre, double-blind, randomized, placebo-controlled study conducted in France using 16-hour nicotine patches. The trial randomized 402 women to either nicotine (n = 203) or placebo patches (n = 199). The study was approved by the Ethics Committee of the Pitié-Salpêtrière Hospital, Paris, France.

Participants

Participants were eligible for inclusion in the SNIPP trial if they smoked at least five cigarettes per day, were aged more than 18 years, of 12–20 weeks' gestation and scored at least 5 on a scale measuring motivation to stop smoking (range 0–10) [17]. Prior to enrolment, participants attended a baseline visit where demographic, obstetric, physiological characteristics and smoking behaviour data were collected, and saliva cotinine concentrations were determined. At this stage, participants were given 2 weeks to quit smoking or reduce the number of cigarettes to fewer than five a day. If after this 2-week period they were unable to do either of these they could be randomized, receive the study drug and set a quit date when treatment began. Participants were asked to stop smoking on a pre-defined quit date and were randomized to either placebo or nicotine patches. Participants were told that they could continue using nicotine patches during smoking lapses. Moreover, patch doses were adjusted according to the previous saliva cotinine determination to optimize the nicotine substitution; this resulted in participants receiving a mean nicotine dose of 18 mg/day [standard deviation (SD) = 6.8] in the nicotine patch arm.

Measures

In the SNIPP trial, abstinence was defined as self-reported abstinence, confirmed by expired air carbon monoxide (CO) concentration ≤ 8 parts per million (p.p.m.) (Smokeanalyzer®; Bedfont Scientific Ltd, Rochester, Kent, UK) [18]. Saliva cotinine samples were collected by placing a cotton roll in the gingival cleft for 1 minute, which was then placed immediately into a Salivette tube (Sarstedt, Nümbrecht, Germany) [18]. Samples were kept at 4°C and were sent to the central biochemistry laboratory (Hôpital Pitié-Salpêtrière, Laboratoire de Biochimie, Dr N. Jacob) within 24 hours for determination [18]. The quantification limit for cotinine was 7.5 mg/l and the between-run coefficient of variation 5–8% [18].

Figure 1 shows when trial visits occurred and measurements were made. Saliva cotinine concentrations were determined at baseline, 2 weeks after quit date and 8 weeks after quit date, with nicotine doses adjusted after each of these visits at 4 and 12 weeks after quit date, respectively. Nicotine doses were adjusted using a conversion factor of 0.1. For example, a saliva cotinine concentration of 100 ng/l equated to a prescription of one 10-mg patch [17]. At baseline, body mass index (BMI), gestational age, ethnicity and Fagerström Test for Cigarette Dependence (FTCD) scores were recorded. As well as at baseline, at each visit women reported any smoking in the previous week, validated by expired air CO. Additionally, intensity of craving for tobacco via the French Tobacco Craving Questionnaire, 12 items (FTCQ-12) and the number of cigarettes smoked by the participant in the last week were assessed. The SNIPP trial recorded cigarette consumption in the past week, rather than cigarettes per day, due to large day-to-day fluctuations in cigarette consumption [19,20]. Partner smoking in the previous week was also assessed, as the second-hand smoke exposure is likely to increase cotinine measures. Women were permitted to use nicotine patches

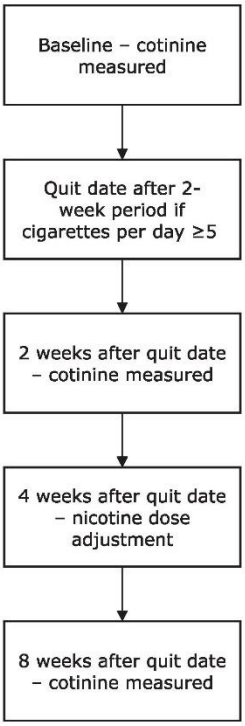

Figure 1 Flow-chart to show each planned visit in the 'Study of Nicotine Patch in Pregnancy' relevant to the current study

from quit date until delivery. A more extensive description is available elsewhere [17].

In this study we used data from women collected at 2 weeks after the quit date who had been allocated nicotine or placebo patches but who reported any smoking in the previous week. A second sample of data collected at 8 weeks after the quit date from women who had smoked in the previous week were used as a sensitivity analysis. Not all women who had cotinine measured at 2 weeks returned for the 8-week visit, and 8-week data also included women who did not return at 2 weeks. We selected women from 2 weeks rather than 8 weeks after the quit date for the main analysis, as this time-point was earlier in gestation, so nicotine metabolism changes since the baseline visit would probably be small and have less impact on findings [21].

Analyses

For baseline data, continuous measures were reported as means with SDs, and categorical measures were reported using frequencies and percentages. Participant and partner's smoking in the previous week were divided by 7 to achieve cigarettes smoked per day. *T*-tests were used to assess whether there were any systematic differences in baseline characteristics between women who were included and those excluded from this study. We used a natural log-transformation of salivary cotinine concentrations to achieve a normal distribution.

For both nicotine and placebo patch groups we used paired *t*-tests to assess 'within-participant' differences between cotinine, CO, cravings, number of cigarettes smoked by the participant and number of cigarettes smoked by their partner, measured at baseline and at 2 weeks. The same analyses were conducted using data from 8 weeks. For saliva cotinine, we present the back-transformed estimates of treatment differences, which is the ratio of the geometric means. Next, we used linear regression analysis to test for an interaction between the measures mentioned above and nicotine patch assignment. We then aimed to identify whether the interactions were significant at increasing increments of baseline values in cotinine, CO, cravings, number of cigarettes smoked by the participant and number of cigarettes smoked by their partner. Findings are presented graphically. *P*-values less than 0.05 were deemed statistically significant. All analyses were conducted using Stata version 15.

After undertaking the planned analyses we generated a Bayes factor from the difference in saliva cotinine, using an online calculator [22]. We were unable to identify any studies that investigated nicotine intake of concurrent smokers and NRT users in pregnancy, so an expected difference of 139.3 ng/ml was taken from a study of nicotine intake outside pregnancy [23]. We used a conservative approach for estimation using a half-normal distribution, where the standard deviation is equal to the expected effect size.

RESULTS

In the SNIPP trial, 203 women were assigned to the nicotine patch arm and 199 women were assigned to the placebo patch arm. At 2 weeks after the quit date, 167 (82.3%) and 148 (74.4%) women returned for the visit in the nicotine patch and placebo patch arms, respectively. In the nicotine patch arm, 149 (73.4%) had smoked in the week prior to the visit and 18 (8.9%) were abstinent whereas, in the placebo group, 131 (65.8%) had smoked during the week prior to the visit and 17 (8.5%) were abstinent. Overall, 12 women had missing cotinine data at this point and were excluded from the study, leaving a sample of

268 for analysis (146 in the nicotine group and 122 in the placebo group).

When comparing SNIPP trial participants excluded from this study with those included, it was found that more women in this study had a partner who smoked. Table 1 gives baseline characteristics of women in both study groups and, using these descriptors, both groups were broadly similar. From the participants who provided 2-week data, those assigned nicotine patch had a mean age of 30 years and gestational age at baseline of 12.8 weeks; therefore, their mean gestational age at 2 weeks post-quit date would be between 16 and 17 weeks.

Table 2 compares indicators of smoking intensity between baseline and 2 weeks after the quit date for pregnant smokers in both the placebo and nicotine patch groups. In the nicotine group there was no significant difference between cotinine concentrations [ratio of geometric means = 0.94 ng/ml, 95% confidence interval (CI) = 0.83–1.07 ng/ml; $P = 0.37$, Bayes factor = 0.15],

Table 1 Participant baseline characteristics; *n* (%) or mean (standard deviation).

Characteristic	*Women on nicotine patch (n = 146)*	*Women on placebo patch (n = 122)*
Age (years)	29.70 (6.00)	28.88 (5.03)
BMI (kg/m^2)	25.52 (5.40)	25.21 (5.33)
Gestational age at baseline (weeks)	12.8 (3.2)	12.6 (5.4)
Ethnicity		
European	139 (95)	115 (94)
African	4 (3)	4 (3)
Asian	1 (1)	1 (1)
Other	2 (1)	2 (2)
Current cigarettes smoking per day		
5–10	66 (45)	55 (45)
11–20	69 (47)	50 (41)
21–30	7 (5)	16 (13)
>30	4 (3)	1 (1)
Fagerström Test for Cigarette Dependence[a]		
Very low	32 (22)	20 (16)
Low	34 (23)	42 (34)
Medium	29 (20)	18 (15)
High	43 (29)	33 (27)
Very high	8 (6)	9 (7)
Partner smoking		
Yes	99 (69)	90 (75)
Saliva cotinine (ng/ml)	143.86 (82.81)	144.36 (74.33)
Expired air carbon monoxide (p.p.m.)	11.8 (6.7)	12.2 (7.3)
French Tobacco Craving Questionnaire score	33.64 (8.60)	35.55 (9.53)

[a]Fagerström Test for Cigarette Dependence is a six-item test where answers are summed to yield a total score of 0–10. The higher the total score, the more intense is the patient's physical dependence on cigarettes; i.e. a score between 0–2 indicates a very low level of dependence on cigarettes, and 8–10 indicates a very high-level dependence on cigarettes [24]. BMI = body mass index; p.p.m. = parts per million.

Table 2 Baseline to 2 weeks after the quit date 'within-participant' differences in indicators of smoking intensity in pregnant smokers by treatment group, with a significance test for interaction with nicotine patch.

	Nicotine patch (n = 146)				*Placebo patch (n = 122)*				
Characteristic	*Baseline mean (SD)*	*2 weeks after quit date mean (SD)*	*Mean difference (95% CI)*	*P-value*[b]	*Baseline mean (SD)*	*2 weeks after quit date mean (SD)*	*Mean difference (95% CI)*	*P-value*[c]	*Interaction P-value*[b]
Saliva cotinine[a] (ng/ml)	117.83	111.14	0.94 (0.83 to 1.07)	0.370	122.46	83.01	0.68 (0.59 to 0.78)	< 0.001	0.148
Expired air carbon monoxide (p.p.m.)	11.8 (6.7)	8.7 (6.5)	−3.0 (−4.2 to −1.9)	< 0.001	12.2 (7.3)	10.2 (9.1)	−2.0 (−3.8 to −0.2)	0.028	0.498
FTCQ-12[d]	33.75 (8.63)	31.38 (8.06)	−2.38 (−3.88 to −0.87)	0.002	35.84 (9.00)	33.36 (8.57)	−2.49 (−4.37 to −0.60)	0.010	0.317
Number of cigarettes smoked per day	12 (6)	6 (5)	−6 (−7 to −5)	< 0.001	12 (6)	6 (6)	−6 (−7 to −5)	< 0.001	0.046
Number of cigarettes partner smoked per day	17 (9)	15 (7)	−1 (−2 to 0)	0.026	16 (7)	14 (7)	−2 (−3 to −1)	0.003	0.168

Paired *t*-tests were used to compare differences at baseline and 2 weeks after the quit date. A linear model was used to test for an interaction of nicotine patch between baseline and 2 weeks. SD = standard deviation; CI = confidence interval; p.p.m. = parts per million. [b]*P*-value for interaction of nicotine patch with indicators of smoking intensity at baseline compared with at 2 weeks after the quit date. [c]*P*-value for the difference between indicators of smoking intensity between baseline and 2 weeks in the nicotine patch group. [d]*P*-value for the difference between indicators of smoking intensity between baseline and 2 weeks in the placebo patch group. [d]FTCQ –12 = French Tobacco Craving Questionnaire score. [a]Back-transformed saliva cotinine data. Means represent geometric means. Mean difference presented as ratio of geometric means.

but CO concentrations significantly decreased from baseline to 2 weeks after the quit date (mean difference − 3.0 p.p.m., 95% CI = −4.2 to −1.9 p.p.m.; $P < 0.001$), whereas the placebo group exhibited a significant reduction in cotinine (ratio of geometric means = 0.68 ng/ml, 95% CI = 0.59–0.78 ng/ml; $P < 0.001$) as well as a reduction in CO concentration (mean difference = −2.0 p.p.m., 95% CI = −3.8 to −0.2 p.p.m., $P < 0.028$). There were also significantly lower levels of craving, lower numbers of cigarettes smoked in the previous week and women's partners were reported to have smoked fewer cigarettes in both nicotine and placebo patch groups.

Table 2 also reports results for interaction tests between the indicators of smoking intensity and nicotine patch assignment. There was a significant interaction between nicotine patch assignment and a reduction in number of cigarettes smoked ($P = 0.046$). This means that women assigned nicotine patches smoked less at week 2 compared to women assigned placebo patches. Interactions between the remaining indicators of smoking intensity and nicotine patch assignment were not significant. Upon further exploration it was discovered that there was an interaction between nicotine patch assignment and women with higher baseline cotinine concentrations (Fig. 2). Women assigned nicotine patches with baseline saliva cotinine concentrations of approximately 90 ng/ml and above had higher cotinine concentrations at week 2 compared to women assigned placebo patches.

In the sensitivity analysis, the 8-week data showed a similar pattern to the 2-week data (Supporting information, Table S1). There was no significant difference between cotinine concentrations at baseline and 8 weeks in the nicotine patch group (ratio of geometric means = 0.85 ng/ml, 95% CI = 0.71–1.00 ng/ml; $P = 0.055$, Bayes factor = 0.12); however, there were significant reductions for all other indicators of smoking intensity aside from craving score (mean difference = −1.69, 95% CI = −3.58 to 0.20 $P = 0.079$). In women assigned placebo patches, there were significant reductions for all indicators of smoking cessation aside from expired CO concentration (mean difference − 2.4 p.p.m., 95% CI = − 5.0 to 0.3 p.p.m., $P < 0.077$). The interaction tests found no significant interaction for nicotine patch assignment; however, graphical exploration found that there was a significant interaction for nicotine patch assignment and participants who reported smoking between 100 and 250 cigarettes a week at baseline (Supporting information, Fig. S1); in these women, assignment to nicotine patch was associated with having smoked fewer cigarettes in the previous 7 days.

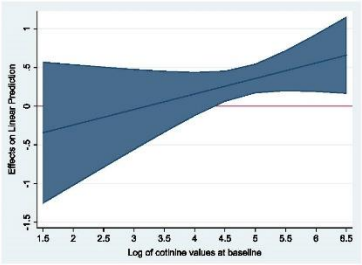

Figure 2 Graph to show interaction of nicotine patches on cotinine concentrations at 2 weeks with increasing baseline cotinine concentrations. The shaded area represents the 95% confidence intervals. As the shaded area for log cotinine > 4.5 is above 0, there is a significant interaction of nicotine patches for an increase in cotinine at 2 weeks in women with log cotinine concentrations of greater than 4.5 (back-transformed to 90 ng/ml), compared with placebo. [Colour figure can be viewed at wileyonlinelibrary.com]

DISCUSSION

Our findings show that women prescribed nicotine patches but who also admitted smoking had similar cotinine concentrations to those generated when they only smoked. These women also reported smoking less and had lower expired air CO readings than when they smoked prior to their quit attempt. In comparison, smokers issued with placebo patches had lower cotinine concentrations than when smoking; they also showed reductions in numbers of cigarettes smoked and expired CO concentrations. Our results also indicate that women who smoke and use nicotine patches smoke less later in pregnancy.

A limitation to our study is that, while we know that women included in this study were prescribed nicotine patches, we have very limited information about how much these were used. However, as study measurements at 2- and 8-week follow up were taken with the intention of personalizing the nicotine doses which women received from patches, it seems very likely that women who attended these appointments were still using these. Furthermore, the SNIPP trial also reports (where adherence data exists) that the median self-reported adherence rate was 85% [17].

Another possible limitation concerns the validity of women's reports of smoking or not smoking during the week prior to having 2- and 8-week measurements taken. In SNIPP, women were defined as smokers if they had reported any smoking in the week prior to a study visit, and this was validated by an expired CO reading. However, expired air CO can only reliably validate smoking status during the previous 6 hours [25] and, although some women may have over- or underestimated the number of cigarettes smoked in the previous week, we could only accurately quantify tobacco smoke exposure in the 6 hours prior to CO measurement. Nevertheless, this could only have had a major impact on findings if women generally

under-reported their smoking during the week prior to follow-up appointments and, in the 6 hours before follow-up appointments, tried to smoke less than they had reported. It seems unlikely that trial participants would do this before attending a nicotine patch dose-titration appointment.

A strength of this study is that the data were obtained as part of a well-conducted randomized controlled trial and included reported smoking behaviour with concurrent CO and cotinine estimation at several time-points. To our knowledge, there has been no previous study that has investigated smoking behaviour and CO exposure from concurrent use of nicotine patches and smoking in pregnancy. Hence, we believe this makes an original contribution to the field. Another strength is that comparisons are based on 'within-participant' measurements; this means that inter-participant variations are very unlikely to explain study findings. Indeed, with this study design one would only expect findings to be affected by characteristics of women which were prone to change between baseline and follow-up. Women's nicotine metabolic rates (NMR) increase as pregnancy progresses, and these would be expected to affect their plasma nicotine concentrations and so, potentially, also their cravings and intensity of smoking [13,21]. However, any effect would seem to be marginal as, even in the placebo group, women reported smoking fewer cigarettes. Also, as pregnancy-related NMR acceleration is generally complete by the end of the first trimester and women's mean gestation at baseline was ~13 weeks, there may have been little scope for this factor to have any influence. It seems likely, therefore, that the differences reported reflect differences in smoking behaviour and not changes in women's physiology during pregnancy.

Our study informs about cotinine concentrations in pregnant women who use nicotine patches but are not abstinent from smoking, and show that cotinine concentrations in such women were no higher than when they were smoking. Additionally, women included in this study had simultaneous and statistically significant reductions in their cigarette use, validated by a reduction in expired CO. This suggests that when pregnant women use nicotine patches and smoke, they smoke less than they would if they were not using nicotine patches. This is important, as it could influence how women are advised to use NRT in pregnancy, i.e. encouraged to continue using NRT despite a relapse.

We are unaware of any previous studies measuring cotinine or CO in smokers who concurrently use NRT during pregnancy. A systematic review and meta-analysis that aimed to identify and describe studies which report nicotine or cotinine concentrations in pregnant women when smoking, and subsequently when abstinent from smoking and using NRT, concluded that among pregnant women who quit smoking, standard-dose NRT generates lower nicotine exposure than smoking [14]. The meta-analysis compared cotinine exposures when pregnant women smoke with those when they use NRT and found that concentrations were, on average, 75.3 ng/ml lower when abstinent and using NRT than when the same women smoked [14]. In SNIPP, salivary cotinine concentrations at baseline (when smoking) were compared to cotinine concentrations at 1 month in women who had stopped smoking but were using nicotine patches. Cotinine concentrations were 98.5 ng/ml while smoking, but only 62.8 ng/ml while using nicotine patches [17]. In our study we found that women who were assigned the placebo patch but admitted to smoking also exhibited reduced cotinine concentrations compared to those when smoking alone.

Most studies in the above review used lower nicotine doses than were used by participants in this paper's analyses; other than SNIPP, studies used standard rather than higher doses of nicotine, and these delivered no more than 15 mg cotinine in 16 hours or the 24-hour equivalent [14]. Thus, when pregnant smokers become abstinent and adhere with such 'standard' doses of NRT they are, on average, exposed to less nicotine than from smoking [14]. In SNIPP, patch doses were adjusted according to the previous saliva cotinine determination to optimize the nicotine substitution leading to somewhat higher mean nicotine doses than usual (18 mg/day, SD = 6.8). It is expected that the dose adjustment would improve nicotine substitution, thus it is possible that women assigned nicotine patches in the 8-week sample would have higher cotinine concentrations than they had at baseline. Despite this adjustment, there was no significant difference in cotinine concentrations in women who were assigned nicotine patches and admitted to smoking compared to those when smoking alone. This also suggests that smoking and using nicotine patches of 'standard' doses may lead to lower cotinine concentrations during pregnancy than smoking alone, prior to pregnancy.

Our findings provide the first data we are aware of which quantifies pregnant women's smoking behaviour when using nicotine patches, and this suggests that when pregnant women use nicotine patches as part of a quit attempt, but also smoke, they smoke less than they did before the quit attempt started. This means that their exposure to the toxic products of burnt tobacco is reduced. A possible reason for this is that women who continue to smoke when using nicotine patches obtain nicotine from both patches and tobacco, and nicotine delivered from patches reduces women's cravings such that they feel less need to 'top up' concentrations of nicotine in their body fluids through smoking. This suggests that clinicians can reassure women that it is alright to smoke and use nicotine patches if, ultimately, they are trying for abstinence.

CONCLUSIONS

In conclusion, despite having similar cotinine exposure to that from cigarette smoking, pregnant women who use nicotine patches and smoke, smoke less and exhale less CO, so their exposure to other tobacco smoke toxins is also likely to be lower.

Declaration of interests

R.C., T.C. and J.L.-B.: none to declare. I.B. received honoraria for consulting and lectures from Pfizer Ltd.

Acknowledgements

R.C.'s research studentship was funded by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care East Midlands (NIHR CLAHRC East Midlands). T.C. is an NIHR Senior Investigator; the views expressed in this article are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. The authors thank Assistance publique-Hôpitaux de Paris for permitting us access to the database used for this analysis.

References

1. Pineles B. L., Park E., Samet J. M. Systematic review and meta-analysis of miscarriage and maternal exposure to tobacco smoke during pregnancy. *Am J Epidemiol* 2014; **179**: 807–23.
2. Flenady V., Wojcieszek A. M., Middleton P., Ellwood D., Erwich J. J., Coory M. *et al.* Stillbirths: recall to action in high-income countries. *Lancet* 2016; **387**: 691–702.
3. Department of Health (DoH) Towards a Smokefree Generation. In: *A Tobacco Control Plan for England*. London: DoH; 2017. (Archived at http://www.webcitation.org/7815xLiiQ on accessed 30 May 2019).
4. McAndrew F., Thompson J., Fellows L., Large A., Speed M., Renfrew M. J. Infant Feeding Survey 2010. Leeds: Health and Social Care Information Centre; 2012. (Archived at http://www.webcitation.org/78l6dgvNw on accessed 30 May 2019).
5. Tong V. T., Dietz P. M., Morrow B., D'Angelo D. V., Farr S. L., Rockhill K. M., et al. Trends in smoking before, during, and after pregnancy—Pregnancy Risk Assessment Monitoring System, United States, 40 sites, 2000–2010. *Morbidity and Mortality Weekly Report: Surveillance Summaries*. 2013;**62**(6):1–19.
6. Blondel B., Kermarrec M. Les naissances en 2010 et leur évolution depuis 2003. Enquête nationale périnatale 2010 [National Perinatal Survey 2010. Births in 2010 and their evolution since 2003]. Paris: Inserm Éditions L'annonce prénatale d'un risque de. handicap 2011. (Archived at http://www.webcitation.org/78l9adPmA on accessed 30 May 2019).
7. National Institute for Health and Clinical Excellence. Public health guidance 26: how to stop smoking in pregnancy and following childbirth. London: 2010 June 2010. Report No.: PH26. (Archived at http://www.webcitation.org/78l88zcUz on accessed 30 May 2019).
8. European Network for Smoking and Tobacco Prevention *European Smoking Cessation Guidelines: The Authoritative Guide to a Comprehensive Understanding of the Implications and Implementation of Treatments and Strategies to Treat Tobacco Dependence*. Belgium: European Network for Smoking and Tobacco Prevention aisbl (ENSP); 2011. (Archived at http://www.webcitation.org/78l9FeWYq on 30 May 2019).
9. Stead L. F., Perera R., Bullen C., Mant D., Hartmann-Boyce J., Cahill K. *et al.* Nicotine replacement therapy for smoking cessation. *Cochrane Database Syst Rev* 2012; **11**: CD000146.
10. Coleman T., Chamberlain C., Davey M.-A., Cooper Sue E., Leonardi-Bee J. Pharmacological interventions for promoting smoking cessation during pregnancy. *Cochrane Database Syst Rev*CD010078.
11. Dempsey D., Tutka P., Jacob P., Allen F., Schoedel K., Tyndale R. F. *et al.* Nicotine metabolite ratio as an index of cytochrome P450 2A6 metabolic activity. *Clin Pharmacol Ther* 2004; **76**: 64–72.
12. Rubinstein M. L., Benowitz N. L., Auerback G. M., Moscicki A.-B. Rate of nicotine metabolism and withdrawal symptoms in adolescent light smokers. *Pediatrics* 2008; **122**: e643–e647.
13. Dempsey D., Jacob P., Benowitz N. L. Accelerated metabolism of nicotine and cotinine in pregnant smokers. *J Pharmacol Exp Ther* 2002; **301**: 594–8.
14. Hickson C., Lewis S., Campbell K. A., Cooper S., Berlin I., Claire R. *et al.* Comparison of nicotine exposure during pregnancy when smoking and abstinent with nicotine replacement therapy: systematic review and meta-analysis. *Addiction* 2018; https://doi.org/10.1111/add.14473
15. Coleman T., Cooper S., Thornton J. G., Grainge M. J., Watts K., Britton J. *et al.* A randomized trial of nicotine-replacement therapy patches in pregnancy. *N Engl J Med* 2012; **366**: 808–18.
16. Ferguson S. G., Gitchell J. G., Shiffman S. Continuing to wear nicotine patches after smoking lapses promotes recovery of abstinence. *Addiction* 2012; **107**: 1349–53.
17. Berlin I., Grangé G., Jacob N., Tanguy M. L. Nicotine patches in pregnant smokers: randomised, placebo controlled, multicentre trial of efficacy. *BMJ* 2014; **348**: g1622.
18. Berlin I., Singleton E. G., Heishman S. J. A comparison of the Fagerström test for cigarette dependence and cigarette dependence scale in a treatment-seeking sample of pregnant smokers. *Nicotine Tob Res* 2015; **18**: 477–83.
19. Pickett K. E., Wakschlag L. S., Dai L., Leventhal B. L. Fluctuations of maternal smoking during pregnancy. *Obstet Gynecol* 2003; **101**: 140–7.
20. Pickett K. E., Rathouz P. J., Kasza K., Wakschlag L. S., Wright R. Self-reported smoking, cotinine levels, and patterns of smoking in pregnancy. *Paediatr Perinat Epidemiol* 2005; **19**: 368–76.
21. Bowker K., Lewis S., Coleman T., Cooper S. Changes in the rate of nicotine metabolism across pregnancy: a longitudinal study. *Addiction* 2015; **110**: 1827–32.
22. Dienes Z. Make the most of your data with Bayes. Available at: http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/Bayes.htm (accessed 8 May 2019) (Archived at http://www.webcitation.org/78DXzcw9f).
23. Shahab L., Beard E., Brown J., West R. Prevalence of NRT use and associated nicotine intake in smokers, recent

ex-smokers and longer-term ex-smokers. *PLOS ONE* 2014; **9**: e113045.
24. Heatherton T. F., Kozlowski L. T., Frecker R. C., Fagerstrom K. O. The Fagerström test for nicotine dependence: a revision of the Fagerstrom tolerance questionnaire. *Br J Addict* 1991; **86**: 1119–27.
25. Benowitz N. L., Jacob P. III, Ahijevych K., Jarvis M. J., Hall S., LeHouezec J. *et al*. Biochemical verification of tobacco use and cessation. *Nicotine Tob Res* 2002; **4**: 149–59.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1 Baseline to 8-weeks after the quit date 'within-participant' differences in indicators of smoking intensity in pregnant smokers by treatment group, with a significance test for interaction with nicotine patch

Figure S1 Graph to show interaction of nicotine patches on cigarettes smoked at 2-weeks with increasing number of cigarettes smoked at baseline. The shaded area represents the 95% confidence intervals. As the shaded area for number of cigarettes smoked between 100–250, is below 0, there is a significant interaction of nicotine patches for a reduction of cigarettes smoked at 8-weeks in women that smoked between 100–250 cigarettes in the week prior to baseline compared with placebo.

Pharmacological interventions for promoting smoking cessation during pregnancy – PROSPERO Record

Systematic review

1. * Review title.

Give the title of the review in English

Pharmacological interventions for promoting smoking cessation during pregnancy

2. Original language title.

For reviews in languages other than English, give the title in the original language. This will be displayed with the English language title.

3. * Anticipated or actual start date.

Give the date the systematic review started or is expected to start.

21/05/2019

4. * Anticipated completion date.

Give the date by which the review is expected to be completed.

30/09/2019

5. * Stage of review at time of this submission.

Tick the boxes to show which review tasks have been started and which have been completed. Update this field each time any amendments are made to a published record.

Reviews that have started data extraction (at the time of initial submission) are not eligible for inclusion in PROSPERO. If there is later evidence that incorrect status and/or completion date has been supplied, the published PROSPERO record will be marked as retracted.

This field uses answers to initial screening questions. It cannot be edited until after registration.

The review has not yet started: No

Review stage	Started	Completed
Preliminary searches	Yes	Yes
Piloting of the study selection process	Yes	Yes
Formal screening of search results against eligibility criteria	Yes	Yes
Data extraction	Yes	Yes
Risk of bias (quality) assessment	Yes	Yes
Data analysis	Yes	No

Provide any other relevant information about the stage of the review here.

6. * Named contact.

The named contact is the guarantor for the accuracy of the information in the register record. This may be any member of the review team.

Ravinder Claire

Email salutation (e.g. "Dr Smith" or "Joanne") for correspondence:

Mr Claire

7. * Named contact email.

Give the electronic email address of the named contact.

ravinder.claire@nottingham.ac.uk

8. Named contact address

Give the full institutional/organisational postal address for the named contact.

Division of Primary Care, School of Medicine,

University of Nottingham,

Room 1502 Tower Building,

University Park,

Nottingham,

NG7 2RD

9. Named contact phone number.

Give the telephone number for the named contact, including international dialling code.

+44 (0)1157486682

10. * Organisational affiliation of the review.

Full title of the organisational affiliations for this review and website address if available. This field may be completed as 'None' if the review is not affiliated to any organisation.

University of Nottingham

Organisation web address:

11. * Review team members and their organisational affiliations.

Give the personal details and the organisational affiliations of each member of the review team. Affiliation refers to groups or organisations to which review team members belong. **NOTE: email and country now MUST be entered for each person, unless you are amending a published record.**

Mr Ravinder Claire. University of Nottingham
Dr Sue Cooper. University of Nottingham
Professor Jo Leonardi-Bee. University of Nottingham
Professor Tim Coleman. University of Nottingham
Dr Mary-Ann Davey. Monash University
Dr Catherine Chamberlain. La Trobe University
Dr Ivan Berlin. Hôpital Pitié?Salpêtrière

12. * Funding sources/sponsors.

Details of the individuals, organizations, groups, companies or other legal entities who have funded or sponsored the review.

Ravinder Claire's research studentship was funded by University of Nottingham in association with the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care East Midlands (NIHR CLAHRC East Midlands). This article presents independent research funded by the National Institute for Health Research (NIHR) under the Programme Grants for Applied Research programme RP-PG-0615-20003. Professor Coleman is an NIHR Senior Investigator; the views expressed in this article are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care

Grant number(s)

State the funder, grant or award number and the date of award

13. * Conflicts of interest.

List actual or perceived conflicts of interest (financial or academic).

Yes

Berlin, Ivan: Ivan Berlin declares occasional honoraria from Pfizer Ltd for consultancy, participation in board meetings and lectures. Receipt of a Australian National Health and Medical Research Council (NHMRC) Career Development Fellowship to support work around life-course approaches to improving health equity in the perinatal period for Aboriginal parents. Receives an NHMRC grant to co-design perinatal strategies to support Aboriginal parents experiencing complex trauma. Contact/lead author for a Cochrane Review entitled 'Psychosocial Interventions to promote smoking cessation in pregnancy' and co-author on an editorial calling for Australian researchers to oppose tobacco industry funding for smoking research.

Claire, Ravinder: none known.

Coleman, Tim: none known within the previous 36 months.

Cooper, Sue E: Sue Cooper is a co-applicant and is employed by funding for an NIHR Programme Grant for Applied Research that includes conducting this review. She was previously employed by funding for the

SNAP (Smoking Nicotine and Pregnancy) Trial, which is included in this review.

Davey, Mary-Ann: none known.

Leonardi-Bee, Jo: Jo Leonardi-Bee reports personal fees from undertaking independent statistical review for Danone Nutricia Research, and a grant from the Food Standards Agency, both outside the submitted work.

14. Collaborators.

Give the name and affiliation of any individuals or organisations who are working on the review but who are not listed as review team members. **NOTE: email and country must be completed for each person, unless you are amending a published record.**

15. * Review question.

State the review question(s) clearly and precisely. It may be appropriate to break very broad questions down into a series of related more specific questions. Questions may be framed or refined using PI(E)COS or similar where relevant.

1. Are pharmacological interventions for smoking cessation used in pregnancy effective for smoking cessation in later pregnancy and after childbirth?

2. Do pharmacological interventions for smoking cessation used in pregnancy affect adverse pregnancy and birth outcomes?

This is an update to a previous Cochrane review. The trial sequential analysis will be performed and published separate to the Cochrane review.

Coleman T, Chamberlain C, Davey MA, Cooper SE, Leonardi?Bee J. Pharmacological interventions for promoting smoking cessation during pregnancy. Cochrane Database of Systematic Reviews 2015, Issue 12. Art. No.: CD010078. DOI: 10.1002/14651858.CD010078.pub2.

16. * Searches.

State the sources that will be searched (e.g. Medline). Give the search dates, and any restrictions (e.g. language or publication date). Do NOT enter the full search strategy (it may be provided as a link or attachment below.)

We will search the Cochrane Pregnancy and Childbirth Group's Trials Register by contacting the Trials Search Co-ordinator using the terms listed in Appendix 1.

The Cochrane Pregnancy and Childbirth Group's Trials Register is maintained by the Trials Search Co-ordinator and contains trials identified from:

1. Monthly searches of the Cochrane Central Register of Controlled Trials (CENTRAL);
2. Weekly searches of MEDLINE (Ovid);
3. Weekly searches of Embase (Ovid);
4. Monthly searches of CINAHL (EBSCO);
5. Hand searches of 30 journals and the proceedings of major conferences;
6. Weekly current awareness alerts for a further 44 journals plus monthly BioMed Central email alerts.

Details of the search strategies for the Cochrane Central Register of Controlled Trials (CENTRAL),

MEDLINE, Embase and CINAHL, the list of hand searched journals and conference proceedings, and the list of journals reviewed via the current awareness service can be found in the 'Specialized Register' section within the editorial information about the Cochrane Pregnancy and Childbirth Group.

The search dates are from 11 July 2015, when the search for the previous review took place, to present day.

17. URL to search strategy.

Upload a file with your search strategy, or an example of a search strategy for a specific database, (including the keywords) in pdf or word format. In doing so you are consenting to the file being made publicly accessible. Or provide a URL or link to the strategy. Do NOT provide links to your search **results**.

Alternatively, upload your search strategy to CRD in pdf format. Please note that by doing so you are consenting to the file being made publicly accessible.

Do not make this file publicly available until the review is complete

18. * Condition or domain being studied.

Give a short description of the disease, condition or healthcare domain being studied in your systematic review.

Tobacco smoking during pregnancy is one of the most significant, potentially-preventable cause of a range of adverse pregnancy outcomes, including placental abruption, stillbirth, miscarriage, preterm birth (less than 37 weeks' gestation) and low birthweight (less than 2500 g), probably through a reduction in the supply of oxygen and other essential fetal nutrients; it is also associated with poorer fetal neuro-development.

Preterm birth is the leading cause of neonatal mortality and morbidity, with up to half of all paediatric neurodevelopmental problems ascribed to preterm birth.

Low birthweight is a surrogate measure of the harmful impact of tobacco smoking on fetal development, and there is evidence of an association between low birthweight and adult morbidities, including coronary heart disease, type 2 diabetes mellitus, and obesity.

19. * Participants/population.

Specify the participants or populations being studied in the review. The preferred format includes details of both inclusion and exclusion criteria.

Inclusion: any women who are pregnant, and are defined as smokers.

Exclusion: non-pregnant women.

20. * Intervention(s), exposure(s).

Give full and clear descriptions or definitions of the interventions or the exposures to be reviewed. The preferred format includes details of both inclusion and exclusion criteria.

Pharmacological treatments aimed at promoting smoking cessation including, but not exclusive to,

treatments that have been proven effective in non-pregnant adults, bupropion, varenicline; and electronic

Trials may also provide text (SMS), ideally tailored to individuals' smoking cessation support or cognitive behaviour therapy (CBT) to participants in active drug and comparator trial arms.

21. * Comparator(s)/control.

Where relevant, give details of the alternatives against which the intervention/exposure will be compared (e.g. another intervention or a non-exposed control group). The preferred format includes details of both inclusion and exclusion criteria.

2. Behavioural support or CBT not attributed to the intervention delivered alongside NRT; or

22. * Types of study to be included.

Give details of the study designs (e.g. RCT) that are eligible for inclusion in the review. The preferred format includes both inclusion and exclusion criteria. If there are no restrictions on the types of study, this should be stated.

Inclusion: randomised controlled trials (RCTs), parallel- or cluster-randomised trials will be eligible for inclusion.

Exclusion: quasi-randomised, cross-over and within-participant trial designs, due to the potential biases associated with these designs.

23. Context.

Give summary details of the setting or other relevant characteristics, which help define the inclusion or exclusion criteria.

24. * Main outcome(s).

Give the pre-specified main (most important) outcomes of the review, including details of how the outcome is defined and measured and when these measurement are made, if these are part of the review inclusion criteria.

Self-reported abstinence from smoking at the latest time point in pregnancy at which this is measured and, where available, with biochemical validation using measures such as exhaled carbon monoxide, saliva cotinine, validated urinary metabolites (e.g. preferred NRT co-efficients), anabasine.

Where available, we will also use prolonged, continuous abstinence measures timed from a quit date set in early pregnancy and which allow temporary lapses to smoking as per the Russell Criteria for outcome measurement in cessation studies. However, point prevalence abstinence measures will be substituted for these as required.

* Measures of effect

Please specify the effect measure(s) e.g. relative risks, odds ratios, risk difference, and/or 'number needed to treat.

At the latest time point in pregnancy at which this is measured.

25. * Additional outcome(s).

List the pre-specified additional outcomes of the review, with a similar level of detail to that required for main outcomes. Where there are no additional outcomes please state 'None' or 'Not applicable' as appropriate

to the review

2. Safety

a. Miscarriage/spontaneous abortion;

b. Stillbirth;

c. Mean unadjusted birthweight;

d. Low birthweight (less than 2500 g);

e. Preterm birth (less than 37 weeks' gestation);

f. Neonatal intensive care unit admissions;

g. Neonatal death;

h. Caesarean section;

i. Congenital anomaly;

j. Maternal hypertension;

k. Infant respiratory symptoms;

l. Infant development.

3. Adherence data.

4. Other adverse effects (serious adverse event data contributed to 'safety' outcomes, above).

5. Any reported long-term adverse effects of smoking cessation pharmacotherapies.

NB: A specific search will not be made for 3 and 4 above but, if present, these data will be extracted from included studies and described qualitatively.

* Measures of effect

Please specify the effect measure(s) for you additional outcome(s) e.g. relative risks, odds ratios, risk difference, and/or 'number needed to treat.

Adherence data will be measured during pregnancy.

26. * Data extraction (selection and coding).

Describe how studies will be selected for inclusion. State what data will be extracted or obtained. State how this will be done and recorded.

Teh studies retrieved during the searches will be screened for relevance, and those identified as being potentially eligible will be fully assessed against the inclusion/exclusion criteria, and decisions made regarding selection/rejection, as appropriate.

We will use the data extraction form that was used for the previous Cochrane Review, and for eligible studies, two review authors (RC and another co-author) will use this to independently extract data from new trials for inclusion. The main data to be extracted will include: eligibility criteria (including: study design, participants), participant characteristics (including: age, marital status, education level, parity, number of cigarettes smoked per day prior to pregnancy), intervention information (including: type of intervention, level of behavioural support,

comparison), and outcome measures (including: biochemically validated/reported cessation later in pregnancy, birth outcomes).

The extracted data will be compared, with any discrepancies being resolved through discussion or, if required, by consulting another co-author. If information regarding any of the details is unclear, we will contact authors of the reports to provide further details. One author (RC) will enter the information into Review Manager Software, with another author (JLB) double checking this for accuracy.

27. * Risk of bias (quality) assessment.

State which characteristics of the studies will be assessed and/or any formal risk of bias/quality assessment tools that will be used.

Two review authors (RC and another co-author) will independently assess the risk of bias for each study using the criteria established in the Cochrane Handbook for Systematic Reviews of Interventions. For all studies, we will assess the following characteristics: random sequence generation; allocation concealment; detection bias; performance bias; incomplete outcome data; biochemical validation of smoking status at primary outcome point; and other sources of bias. We will judge all domains, using one of the following three categories: "low risk of bias" "unclear risk of bias" or "high risk of bias", according to the guidelines proposed by the Cochrane Handbook for Systematic Reviews of Interventions. We will resolve any discrepancies through discussion with a third co-author.

28. * Strategy for data synthesis.

Describe the methods you plan to use to synthesise data. This **must not be generic text** but should be **specific to your review** and describe how the proposed approach will be applied to your data. If meta-analysis is planned, describe the models to be used, methods to explore statistical heterogeneity, and software package to be used.

We will carry out the statistical analysis using Review Manager software. Following the standard methods of the Cochrane Tobacco Addiction Group for pharmacological interventions we will use a fixed-effects model for meta-analyses of smoking abstinence data. For meta-analyses of safety and adverse events data we used random-effects models as effects are likely to vary across populations due to significant differences in baseline risk. Results will be presented as risk ratios with 95% confidence intervals.

If there are at least 10 studies included in the meta-analyses, we will consider performing random effects meta-regression analyses to further explore reasons for heterogeneity, using Stata software. A caveat to using this method for adherence data is that there is currently no standard method for reporting adherence; however, for a meta-regression to be undertaken, studies must report adherence data similarly.

We will perform trial sequential analysis (TSA), as this methodology takes into account the volume of significance testing which has been undertaken and adjusts the thresholds that are used to define whether or not results are considered statistically significant. We will add the trials according to the year of publication,

and if more than one trial has been published in a year, we will add trials alphabetically according to the last name of the first author. We plan to estimate the required information size based on a type I error of 5% and a type II error of 10% to detect a risk ratio observed in trials with a low risk of bias.

We will also perform a sensitivity analysis using the parameters above, but using the upper bound of the 95% confidence interval for heterogeneity calculated by the TSA software. This TSA section will not be part of the Cochrane review, and will be published separately.

29. * Analysis of subgroups or subsets.

State any planned investigation of 'subgroups'. Be clear and specific about which type of study or participant will be included in each group or covariate investigated. State the planned analytic approach.

For smoking cessation outcomes, we will explore reasons for heterogeneity between the studies using subgroup analyses based on the following groups:

1. Placebo controlled RCTs;

2. Studies using different types of NRT both alone and in combination (i.e., short- and long-acting NRT);

3. Low dose NRT (less than 10mg/24hr) vs high dose NRT (greater than 10mg/24hr).

If there are at least 10 studies included in the meta-analyses, we will consider performing random effects meta-regression analyses to further explore reasons for heterogeneity, and will also perform a sensitivity analysis using the parameters outlined above.

30. * Type and method of review.

Select the type of review, review method and health area from the lists below.

Type of review

Cost effectiveness
No

Diagnostic
No

Epidemiologic
No

Individual patient data (IPD) meta-analysis
No

Intervention
Yes

Meta-analysis
Yes

Methodology
No

Narrative synthesis

No

Network meta-analysis
No

Pre-clinical
No

Prevention
No

Prognostic
No

Prospective meta-analysis (PMA)
No

Review of reviews
No

Service delivery
No

Synthesis of qualitative studies
No

Systematic review
Yes

Other
Yes

Trial sequential analysis

Health area of the review

Alcohol/substance misuse/abuse
Yes

Blood and immune system
No

Cancer
No

Cardiovascular
No

Care of the elderly
No

Child health
No

Complementary therapies
No

COVID-19
No

Crime and justice
No

Dental

No

Digestive system
No

Ear, nose and throat
No

Education
No

Endocrine and metabolic disorders
No

Eye disorders
No

General interest
No

Genetics
No

Health inequalities/health equity
No

Infections and infestations
No

International development
No

Mental health and behavioural conditions
No

Musculoskeletal
No

Neurological
No

Nursing
No

Obstetrics and gynaecology
No

Oral health
No

Palliative care
No

Perioperative care
No

Physiotherapy
No

Pregnancy and childbirth
Yes

Public health (including social determinants of health)
No

Rehabilitation
No

Respiratory disorders
No

Service delivery
No

Skin disorders
No

Social care
No

Surgery
No

Tropical Medicine
No

Urological
No

Wounds, injuries and accidents
No

Violence and abuse
No

31. Language.

Select each language individually to add it to the list below, use the bin icon to remove any added in error.
English

There is not an English language summary

32. * Country.

Select the country in which the review is being carried out. For multi-national collaborations select all the countries involved.

Australia
England
France

33. Other registration details.

Name any other organisation where the systematic review title or protocol is registered (e.g. Campbell, or The Joanna Briggs Institute) together with any unique identification number assigned by them. If extracted data will be stored and made available through a repository such as the Systematic Review Data Repository (SRDR), details and a link should be included here. If none, leave blank.

34. Reference and/or URL for published protocol.

If the protocol for this review is published provide details (authors, title and journal details, preferably in Vancouver format)

Add web link to the published protocol.

Or, upload your published protocol here in pdf format. Note that the upload will be publicly accessible.

No I do not make this file publicly available until the review is complete

Please note that the information required in the PROSPERO registration form must be completed in full even if access to a protocol is given.

35. Dissemination plans.

Do you intend to publish the review on completion?

Yes

Give brief details of plans for communicating review findings.?

36. Keywords.

Give words or phrases that best describe the review. Separate keywords with a semicolon or new line. Keywords help PROSPERO users find your review (keywords do not appear in the public record but are included in searches). Be as specific and precise as possible. Avoid acronyms and abbreviations unless these are in wide use.

Tobacco Use Cessation Devices

Nicotinic Agonists

Nicotine Replacement Therapy

Varenicline

Pregnancy

Pregnancy Complications

Pregnancy Outcome

Smoking Cessation

37. Details of any existing review of the same topic by the same authors.

If you are registering an update of an existing review give details of the earlier versions and include a full bibliographic reference, if available.

This is an update of the following review: Coleman T, Chamberlain C, Davey MA, Cooper SE, Leonardi?Bee J. Pharmacological interventions for promoting smoking cessation during pregnancy. Cochrane Database of Systematic Reviews 2015, Issue 12. Art. No.: CD010078. DOI: 10.1002/14651858.CD010078.pub2.

38. * Current review status.

Update review status when the review is completed and when it is published.New registrations must be ongoing.
Please provide anticipated publication date

Review_Ongoing

39. Any additional information.

Provide any other information relevant to the registration of this review.

40. Details of final report/publication(s) or preprints if available.

Leave empty until publication details are available OR you have a link to a preprint. List authors, title and journal details preferably in Vancouver format.

Give the link to the published review or preprint.

Pharmacological interventions for promoting smoking cessation during pregnancy – Cochrane Review (Summary)

Cochrane Library

Cochrane Database of Systematic Reviews

Pharmacological interventions for promoting smoking cessation during pregnancy (Review)

Claire R, Chamberlain C, Davey MA, Cooper SE, Berlin I, Leonardi-Bee J, Coleman T

Claire R, Chamberlain C, Davey MA, Cooper SE, Berlin I, Leonardi-Bee J, Coleman T.
Pharmacological interventions for promoting smoking cessation during pregnancy.
Cochrane Database of Systematic Reviews 2020, Issue 3. Art. No.: CD010078.
DOI: 10.1002/14651858.CD010078.pub3.

www.cochranelibrary.com

[Intervention Review]

Pharmacological interventions for promoting smoking cessation during pregnancy

Ravinder Claire[1], Catherine Chamberlain[2], Mary-Ann Davey[3], Sue E Cooper[1], Ivan Berlin[4], Jo Leonardi-Bee[5], Tim Coleman[1]

[1]Division of Primary Care, University of Nottingham, Nottingham, UK. [2]Judith Lumley Centre, La Trobe University, Melbourne, Australia. [3]Department of Obstetrics and Gynaecology, Monash University, Clayton, Australia. [4]Department of Pharmacology, Sorbonne Université, Faculté de medicine-Hopital Pitie-Salpetriere, Paris, France. [5]Centre for Evidence Based Healthcare, Division of Epidemiology and Public Health, Clinical Sciences Building Phase 2, University of Nottingham, Nottingham, UK

Contact address: Tim Coleman, Division of Primary Care, University of Nottingham, Room 1502, Tower Building, University Park, Nottingham, Nottinghamshire, NG7 2RD, UK. tim.coleman@nottingham.ac.uk.

Editorial group: Cochrane Tobacco Addiction Group.
Publication status and date: New search for studies and content updated (no change to conclusions), published in Issue 3, 2020.

Citation: Claire R, Chamberlain C, Davey MA, Cooper SE, Berlin I, Leonardi-Bee J, Coleman T. Pharmacological interventions for promoting smoking cessation during pregnancy. *Cochrane Database of Systematic Reviews* 2020, Issue 3. Art. No.: CD010078. DOI: 10.1002/14651858.CD010078.pub3.

ABSTRACT

Background

Tobacco smoking in pregnancy causes serious health problems for the developing fetus and mother. When used by non-pregnant smokers, pharmacotherapies (nicotine replacement therapy (NRT), bupropion, and varenicline) are effective for increasing smoking cessation, however their efficacy and safety in pregnancy remains unknown. Electronic cigarettes (ECs) are becoming widely used, but their efficacy and safety when used for smoking cessation in pregnancy are also unknown.

Objectives

To determine the efficacy and safety of smoking cessation pharmacotherapies and ECs used during pregnancy for smoking cessation in later pregnancy and after childbirth, and to determine adherence to smoking cessation pharmacotherapies and ECs for smoking cessation during pregnancy.

Search methods

We searched the Cochrane Pregnancy and Childbirth Group's Trials Register (20 May 2019), trial registers, and grey literature, and checked references of retrieved studies.

Selection criteria

Randomised controlled trials (RCTs) conducted in pregnant women, comparing smoking cessation pharmacotherapy or EC use with either placebo or no pharmacotherapy/EC control. We excluded quasi-randomised, cross-over, and within-participant designs, and RCTs with additional intervention components not matched between trial arms.

Data collection and analysis

We followed standard Cochrane methods. The primary efficacy outcome was smoking cessation in later pregnancy; safety was assessed by 11 outcomes (principally birth outcomes) that indicated neonatal and infant well-being. We also collated data on adherence to trial treatments. We calculated the risk ratio (RR) or mean difference (MD) and the 95% confidence intervals (CI) for each outcome for each study, where possible. We grouped eligible studies according to the type of comparison. We carried out meta-analyses where appropriate.

Main results

We included 11 trials that enrolled a total of 2412 pregnant women who smoked at enrolment, nine trials of NRT and two trials of bupropion as adjuncts to behavioural support, with comparable behavioural support provided in the control arms. No trials investigated varenicline or ECs. We assessed four trials as at low risk of bias overall. The overall certainty of the evidence was low across outcomes and comparisons as assessed using GRADE, with reductions in confidence due to risk of bias, imprecision, and inconsistency.

Compared to placebo and non-placebo (behavioural support only) controls, there was low-certainty evidence that NRT increased the likelihood of smoking abstinence in later pregnancy (RR 1.37, 95% CI 1.08 to 1.74; I^2 = 34%, 9 studies, 2336 women). However, in subgroup analysis by comparator type, there was a subgroup difference between placebo-controlled and non-placebo controlled RCTs (test for subgroup differences P = 0.008). There was unclear evidence of an effect in placebo-controlled RCTs (RR 1.21, 95% CI 0.95 to 1.55; I^2 = 0%, 6 studies, 2063 women), whereas non-placebo-controlled trials showed clearer evidence of a benefit (RR 8.55, 95% CI 2.05 to 35.71; I^2 = 0%, 3 studies, 273 women). An additional subgroup analysis in which studies were grouped by the type of NRT used found no difference in the effectiveness of NRT in those using patches or fast-acting NRT (test for subgroup differences P = 0.08).

There was no evidence of a difference between NRT and control groups in rates of miscarriage, stillbirth, premature birth, birthweight, low birthweight, admissions to neonatal intensive care, caesarean section, congenital abnormalities, or neonatal death. In one study infants born to women who had been randomised to NRT had higher rates of 'survival without developmental impairment' at two years of age compared to the placebo group. Non-serious adverse effects observed with NRT included headache, nausea, and local reactions (e.g. skin irritation from patches or foul taste from gum), but data could not be pooled. Adherence to NRT treatment regimens was generally low.

We identified low-certainty evidence that there was no difference in smoking abstinence rates observed in later pregnancy in women using bupropion when compared to placebo control (RR 0.74, 95% CI 0.21 to 2.64; I^2 = 0%, 2 studies, 76 women). Evidence investigating the safety outcomes of bupropion use was sparse, but the existing evidence showed no difference between the bupropion and control group.

Authors' conclusions

NRT used for smoking cessation in pregnancy may increase smoking cessation rates in late pregnancy. However, this evidence is of low certainty, as the effect was not evident when potentially biased, non-placebo-controlled RCTs were excluded from the analysis. Future studies may therefore change this conclusion. We found no evidence that NRT has either positive or negative impacts on birth outcomes; however, the evidence for some of these outcomes was also judged to be of low certainty due to imprecision and inconsistency. We found no evidence that bupropion may be an effective aid for smoking cessation during pregnancy, and there was little evidence evaluating its safety in this population. Further research evidence on the efficacy and safety of pharmacotherapy and EC use for smoking cessation in pregnancy is needed, ideally from placebo-controlled RCTs that achieve higher adherence rates and that monitor infants' outcomes into childhood. Future RCTs of NRT should investigate higher doses than those tested in the studies included in this review.

PLAIN LANGUAGE SUMMARY

Drug treatments and electronic cigarettes for stopping smoking in pregnancy

What is the issue?

Smoking during pregnancy harms women and infants. However, many women who smoke struggle to stop whilst pregnant. Medication for smoking cessation reduces the intensity of cravings, meaning that people trying to stop smoking are more likely to succeed in the long term. Providing pregnant women who smoke with these treatments could help them to stop smoking and have a positive impact on both their own health and the health of their infants.

Why is this important?

Medications commonly used to help people to stop smoking include nicotine replacement therapy (NRT), bupropion, and varenicline. Electronic cigarettes containing nicotine are also used by some who smoke to help avoid smoking. However, the safety and effectiveness of smoking cessation drugs and electronic cigarettes in pregnant women is unknown. We searched for studies looking at how good these aids were at helping pregnant women stop smoking and how safe they were when used during pregnancy.

What evidence did we find?

We searched for evidence on 20 May 2019 and identified 11 randomised studies (studies in which participants are assigned to one of two or more treatment groups using a random method) that enrolled a total of 2412 women. Nine studies tested NRT used alongside counselling to stop smoking, whilst the other two studies tested bupropion.

Low-quality evidence suggests that NRT combined with behavioural support might help women to stop smoking in later pregnancy more than behavioural support alone. Medication trials often use placebos, that is tablets or patches that look like the drug but do not actually include it, so that each comparison group has equal expectation of success and there is a fairer test of the benefits of the medicine itself. When just the higher-quality, placebo-controlled trials were analysed, the evidence suggested that NRT was more effective than placebo NRT. There was no evidence that either nicotine patches or fast-acting NRT (such as gum or lozenge) was more effective than the other.

Low-quality evidence suggests that bupropion may be no more effective than placebo in helping women quit smoking later in pregnancy. We found no trials investigating other smoking cessation pharmacotherapies or electronic cigarettes.

There was insufficient evidence to conclude whether NRT had either positive or negative impacts on rates of miscarriage, stillbirth, preterm birth (less than 37 weeks), mean birthweight, low birthweight (less than 2500 g), admissions of babies to neonatal intensive care, or newborn deaths. However, in one trial where infants were followed until two years of age, those infants born to women who had been randomised to NRT were more likely to have healthy development. Similarly, it is unclear whether bupropion had a positive or negative impact on birth outcomes.

Studies that looked at whether women used their stop smoking medications as instructed found that use was generally low, and the majority of women used little of the NRT they were given.

What does this mean?

More research evidence is needed, in particular placebo-controlled trials that test higher doses of NRT, encourage women to use sufficient medication, and follow infants into childhood. Furthermore, more studies are required investigating the effect and safety of bupropion, electronic cigarettes, and varenicline for giving up smoking during pregnancy.

Fetal safety of nicotine replacement therapy in pregnancy: systematic review and meta-analysis

ADDICTION
SSA SOCIETY FOR THE STUDY OF ADDICTION
REVIEW
doi:10.1111/add.15185

Fetal safety of nicotine replacement therapy in pregnancy: systematic review and meta-analysis

Lauren Taylor[1], Ravinder Claire[1], Katarzyna Campbell[1], Tom Coleman-Haynes[1], Jo Leonardi-Bee[2], Catherine Chamberlain[3], Ivan Berlin[4], Mary-Ann Davey[5], Sue Cooper[1] & Tim Coleman[1]

Division of Primary Care, Faculty of Medicine and Health Sciences, University of Nottingham, UK,[1] Centre for Evidence Based Healthcare, Division of Epidemiology and Public Health, Clinical Sciences Building Phase 2, University of Nottingham, Nottingham, UK,[2] Judith Lumley Centre, La Trobe University, Melbourne, Australia,[3] Department of Pharmacology, Sorbonne Université, Faculté de médecine-Hôpital Pitié-Salpetrière, Paris, France[4] and Department of Obstetrics and Gynaecology, Monash University, Clayton, Australia[5]

ABSTRACT

Background and aims Smoking in pregnancy causes substantial avoidable harm to mothers and offspring; nicotine replacement therapy (NRT) may prevent this, and is used to help women to quit. A recently updated Cochrane Review of randomized controlled trials (RCTs) investigating impacts of NRT in pregnancy focuses primarily on efficacy data, but also reports adverse impacts from NRT. Here we identify and summarize NRT impacts on adverse pregnancy outcomes reported in non-randomized controlled trials (non-RCTs). **Methods** Systematic reviews and meta-analyses of RCTs and non-RCT studies of NRT in pregnancy, with design-specific risk of bias assessment and grading of recommendations, assessment, development and evaluations (GRADE) criteria applied to selected outcomes. **Findings** Relevant Cochrane Review findings are reported alongside those from this new review. Seven RCTs were included; $n = 2340$. Nine meta-analyses were performed; non-statistically significant estimates indicated potentially reduced risk from NRT compared with smoking for mean birth weight, low birth weight, preterm birth, intensive care admissions, neonatal death, congenital anomalies and caesarean section and potentially increased risks for miscarriage and stillbirth. GRADE assessment for mean birth weight and miscarriage outcomes indicated 'low' confidence in findings. Twenty-three non-RCTs were included; $n = 931163$. Eleven large studies from five routine health-care cohorts reported clinical outcomes; 12 small studies investigated mainly physiological outcomes within in-patient women given NRT. Findings from meta-analyses for congenital anomalies, stillbirth and preterm birth were underpowered and not in a consistent direction; GRADE assessment of confidence in findings was 'very low'. Routine health-care studies were of higher quality, but implications of reported findings were unclear as there was inadequate measurement and reporting of women's smoking. **Conclusions** Available evidence from randomized controlled trials and non-randomized comparative studies does not currently provide clear evidence as to whether maternal use of nicotine replacement therapy during pregnancy is harmful to the fetus.

Keywords Birth outcomes, fetal health, health outcomes, nicotine replacement therapy, pregnancy, smoking, smoking cessation.

Correspondence to: Lauren Taylor, Division of Primary Care, Faculty of Medicine and Health Sciences, University of Nottingham, Tower Building, University Park, Nottingham NG7 2RD, UK. E-mail: lauren.taylor9@nhs.net
Submitted 1 October 2019; initial review completed 2 March 2020; final version accepted 5 June 2020

INTRODUCTION

Smoking in pregnancy has adverse effects on the health of pregnant women and their offspring in the pre- and perinatal periods and in later life [1–3]. Smoking rates are highest among younger, socially disadvantaged pregnant women [4,5], and up to 38% of socio-economic inequalities in stillbirths and infant deaths can be attributed to smoking [6]. Stopping smoking in pregnancy improves birth outcomes [7] and reduces the burden of health-care costs to the National Health Service (NHS) [8].

The National Institute for Health and Care Excellence (NICE) recommends nicotine replacement therapy (NRT) in those women who are unable to stop smoking with non-pharmacological interventions [9]. However, even when pregnant women choose NRT, many do not use this

for very long [10] and adherence to NRT by pregnant women tends to be lower than in non-pregnant smokers [10–12]. This poor adherence may at least partially explain why NRT has been found to be less effective when used in pregnancy [13]. One possible reason for poor adherence to NRT in pregnancy is maternal concern about the safety of NRT. Qualitative interviews with pregnant women who sought support from NHS Stop Smoking Services demonstrated that they often reported using NRT intermittently or stopping courses early due to safety concerns [14].

There is a strong theoretical rationale for using NRT to avoid smoking in pregnancy: even if women do not stop smoking completely, cigarette smoke exposes the fetus to numerous toxins whereas NRT exposes them to only nicotine, and so is very likely to be safer [15]. A Cochrane Review investigating the impacts of NRT in pregnancy has recently been updated [13]. RCTs produce the least biased evidence but they also generally have small sample sizes, such that even when they are combined in meta-analyses, small adverse impacts may not be detected. Well-conducted, large non-RCT studies may be still prone to bias, but comprehensive confounder-adjustment could augment RCT data and provide sufficient power to investigate infrequent health outcomes following NRT use in pregnancy. The Cochrane Review focuses primarily on efficacy data, with adverse effects reported as secondary outcomes. Consequently, we conducted a systematic review of non-RCT studies reporting usually adverse fetal or infant health outcomes after pregnant women's use of NRT. Here we report this process alongside the safety-orientated findings from the updated Cochrane Review [13], with the aim of providing a comprehensive, objective and contemporary assessment of whether and how use of NRT during gestation affects pregnancy outcomes.

METHODS

Randomized controlled studies (RCTs)

Standard Cochrane Review (CR) methods used are described in the published review [13]. Searches, for RCTs only, were concluded by 20 May 2019 and from included studies we extracted data on the following outcomes: miscarriage/spontaneous abortion; stillbirth; birth weight; low birth weight (< 2500 g); preterm birth (< 37 weeks' gestation); neonatal intensive care unit admissions; neonatal death; caesarean section; congenital anomalies; infant development; and respiratory symptoms. We assessed study quality using Cochrane's 'risk of bias' tool. A priori, we planned to use grading of recommendations, assessment, development and evaluations (GRADE) criteria for birth weight and miscarriage/spontaneous abortion outcomes, to report studies separately where in meta-analyses $I^2 > 75\%$, and to conduct subgroup analyses for placebo and non-placebo RCTs.

Non-RCTs

A study protocol, written in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) statement, was registered on PROSPERO (International Prospective Register of Systematic Reviews) [16,17].

Inclusion criteria

We sought published non-RCT studies, of any design, in any language, reporting empirical data on potentially adverse fetal or infant health outcomes following NRT exposure or nicotine administration in pregnancy. Although we wanted to identify all health outcomes, we anticipated a priori that these would include at least some of the important clinical outcomes in a relevant 2015 Cochrane Review [18] (see below).

Exclusion criteria

We excluded RCTs and studies which reported only smoking-cessation outcomes [18].

Search strategy

A search strategy was developed in MEDLINE and then adapted for the CINAHL, Embase, PsycINFO, CAB Abstracts, Social Sciences Citation Index and Economic and Social Research Council databases. Supporting information, Table S1 gives search terms; we combined those relevant to pregnancy and fetal health with those referring to NRT or nicotine use. NRT became available in the 1980s, so we searched between 1980 and 12 June 2020, hand-searching references from retrieved full texts, including references from texts excluded from the review. Authors were contacted, as required, for study details.

Study selection and data extraction

One reviewer screened titles and abstracts, rejecting those which were not eligible for inclusion and retrieving manuscripts which appeared potentially includable or about which there was uncertainty. Two reviewers independently screened the full texts and a consensus decision was made on inclusion; if consensus was not possible, a third reviewer adjudicated. Study data were extracted by one reviewer and checked by a second, using a piloted form within Covidence (web-based systematic review platform) [19]. Extracted data included: author's details, publication date, study design and objectives, recruitment and data collection methods, participants' characteristics and study outcomes. For NRT exposure, we extracted data concerning when women were issued with or reported using this, and how many times and by what method these data were acquired. We also extracted smoking behaviour data, and particularly any information on smoking before

and after NRT use, including how often and by what means, this was recorded.

Quality assessment

Two researchers independently quality-assessed studies using modified versions of the Newcastle–Ottawa scale (NOS) [20]. Disagreements about scoring were discussed and consensus reached using a third assessor, if necessary. One modified scale was created for studies in which NRT was used as part of routine clinical care; this had a maximum score of eight stars. The other was used for smaller cohorts in which NRT was an experimental intervention (maximum score: seven stars). Both assessed three domains: 'selection', 'design and analysis' and 'outcome', and were modified by removal of the 'demonstration that outcome of interest was not present at start of study' item as pregnancy outcomes could only occur at childbirth. The 'comparability' domain was renamed 'design and analysis', and we removed 'was follow up long enough for outcomes to occur?' from the 'outcome' domain. Supporting information, Appendix S1 details scale modifications and scoring.

Meta-analysis and GRADE criteria

We anticipated substantial variation in study designs and outcomes, so decisions about meta-analyses were made only after consideration of all included studies. Where appropriate, we planned to pool data comparing outcomes following NRT exposure with no NRT exposure. To provide contextual information within the same studies we also compared outcomes following reported NRT exposure with those after smoking.

We created three exposure groups; those women who: (i) were prescribed or reported being given or using NRT, (ii) reported smoking but not being given NRT or (iii) neither reported smoking nor using NRT. As the only indication for using NRT in pregnancy is as a substitute for smoking, we assumed that all women issued with NRT would have smoked prior to this, so where studies categorized women as only having used NRT and not having smoked, we combined these groups with NRT-exposed groups from other studies which did not make this claim. Hence, we assumed that all women issued NRT would have smoked at some point in pregnancy. Review Manager version 5 software generated pooled risk ratios (RR) using a random-effects model and an estimate of heterogeneity using the I^2 statistic from the Mantel–Haenszel model [21]. As non-RCTs and RCTs are subject to very different biases and effects from unmeasured confounding, we decided to present non-RCT and RCT studies in separate meta-analyses. We anticipated that confounding due to women's smoking before, during or after use of NRT was likely to be particularly important to estimates derived from meta-analyses of non-RCTs, as few empirical studies attempted to adjust for this.

Table 1 shows GRADE [22] criteria that were applied to assess strength of evidence for each meta-analysed outcome. These rate the quality or certainty of evidence as 'very low', 'low', 'moderate' or 'high quality'; ratings start at 'high quality' for RCTs and 'low quality' for observational studies and GRADE criteria are used to up/downgrade ratings, as appropriate. Two reviewers independently applied criteria for each meta-analysed outcome; disagreements were resolved by consensus [13].

RESULTS

RCTs

Full results, including the PRISMA diagram, are found in the published CR [13], but of nine RCTs which investigated NRT use in pregnancy, seven reported infant and fetal safety outcomes [23–29] and all were conducted in high-income countries (n = 2340). All RCTs recruited pregnant women who smoked and, as with non-RCTs, pregnancies would have been exposed to tobacco smoke before women joined trials. RCT groups all received either behavioural support alone or with a placebo, or active NRT. Four placebo-RCTs were judged to be at low [23,24,26,29] and two non-placebo RCTs at high risk of bias [25,28]; for the remaining study this was unclear [27]. High bias risk was generally allocated to studies with no placebo control.

All seven studies reported mean birth weight and gestational age at delivery and incidences of low birth weight (below 2500 g). Six reported rates of preterm birth (birth before 37 weeks), miscarriage or spontaneous abortion and stillbirth [23,24,26–29] and four reported rates of infants' admissions to special care and of neonatal death [23,24,26,28]. Three trials reported rates of congenital malformation [23,24,27] and two reported caesarean section rates [23,24]. One study [30] reported infants' 'survival without developmental impairment' and respiratory symptoms at 2 years.

Meta-analysis results: RCTs

Figure 1 shows RCT meta-analyses findings. There was no evidence of a difference in risk of miscarriage/spontaneous abortion between NRT and control groups [RR = 1.60, 95% confidence interval (CI) = 0.53–4.83, I^2 = 0%; Fig. 1.1]. Similarly, there was no evidence of a difference between the numbers of stillbirths in the NRT and control groups (RR = 1.24, 95% CI = 0.54 to 2.84, I^2 = 0%; Fig. 1.2). The pooled estimate for birth weight was higher for the NRT than for the control group, but the CIs incorporated a small decrease in birth weight as well as a more substantial increase, and heterogeneity was high [mean

Table 1 GRADE criteria for assessing non-RCTs.

GRADE criteria	*Reasons to downgrade*
Risk of bias	Studies scoring < 6/8 for risk of bias in the quality assessment were reviewed and if perceived to have such a high risk of bias that they could threaten findings' accuracy, downgrading by one level occurred
Inconsistency	If I^2 was > 50%, effect estimates for each study in the meta-analysis were assessed. If they were very different, with little-to-no overlap of the confidence intervals around studies' effect estimates, rating was downgraded by one level
Indirectness	This criterion assesses if evidence included in the review directly answers the review question. Quality of evidence was not downgraded based on this criterion due to the problem/patient/population, intervention/indicator, comparison, outcome (PICO) criteria used when searching. We felt our narrow PICO criteria meant that all studies included were reporting data that answered the review question, as we wanted information on all health outcomes reported after NRT exposure in pregnancy
Imprecision	If the confidence interval for the effect estimate was so wide that it could be consistent with having an effect in either direction, this was deemed to be a sign of imprecision and rating was downgraded by one level
Publication bias	Quality of evidence not downgraded based on this criterion due to the types of studies appraised
Upgrading	Quality of evidence not upgraded as there was no supporting evidence for the three recommended reasons to upgrade: large magnitude of effect, the presence of a dose–response gradient or that the effect of all plausible confounding factors would be to reduce the effect seen. It is also not recommended to upgrade a downgraded outcome

Criteria derived from the grading of recommendations, assessment, development and evaluation (GRADE) Working Group Handbook [22]. For all criteria, meta-analysed studies' quality was judged against reasons to downgrade. If there was serious concern regarding any criteria (except 'upgrading'), quality of evidence was downgraded to 'very low' quality, from the starting level of 'low' for observational (non-randomized controlled trial) studies.

difference (MD) = 99.73 g, 95% CI = −6.65 to 206.10, I^2 = 70%; Fig. 1.3]. There was no evidence of a difference in the incidence of low birth weight and there was much heterogeneity in the analysis (RR = 0.69, 95% CI = 0.39–1.20, I^2 = 69%; Fig. 1.4).

Analyses of rates of preterm births (RR = 0.81, 95% CI = 0.59–1.11, I^2 = 21%; Fig. 1.5), neonatal intensive care unit admissions (RR = 0.90, 95% CI = 0.64–1.27; I^2 = 0%; Fig. 1.6) and neonatal deaths (RR = 0.66, 95% CI = 0.17 to 2.62, I^2 = 0%; Fig. 1.7) all resulted in CIs spanning one, incorporating the potential for both benefit and harm. Similarly, meta-analyses of congenital anomalies and caesarean birth suggested no clear evidence for a benefit or harm from NRT (congenital anomalies: RR = 0.73, 95% CI = 0.36–1.48, I^2 = 0%, Fig. 1.8; caesarean section: RR = 1.18, 95% CI = 0.83–1.69, I^2 = 46%, Fig. 1.9).

GRADE assessment found a 'low' certainty of evidence for mean birth weight and miscarriage/spontaneous abortion outcomes.

Narratively reported outcomes: RCTs

Two RCTs [23,24] reported the distribution of Apgar scores at 5 minutes after birth, cord arterial blood pH, intraventricular haemorrhage, neonatal convulsions, necrotizing enterocolitis, mechanical ventilation of infant, assisted vaginal delivery and maternal death between NRT and placebo groups; no statistically significant differences were noted. One RCT [30] reported infant outcomes after the neonatal period. Using a composite self-report outcome based on the Ages and Stages Questionnaire, 3rd edition instrument [31], significantly better infant developmental outcomes were observed in infants born to women who had been randomized to NRT compared to those in the placebo group. The odds ratio (OR) for infants reaching 2 years of age 'without developmental impairment' (i.e. normal development) was 1.40 (95% CI = 1.05–1.86). However, there was no difference in parental reports of infants' respiratory symptoms; the OR for reporting of any respiratory problem in the NRT group was 1.32 (95% CI = 0.97–1.74).

Non-RCT studies

Study selection, characteristics and outcome measures

A total of 18467 titles and abstracts were identified and, after duplicate removal, 9391 records were screened. Forty-five full text articles were retrieved and 23 were included in the review; Fig. 2 shows the reasons for study exclusion.

Table 2 presents characteristics of the 23 included studies (n = 931 163). Eleven were conducted in health-care settings, used routine clinical data [32–42], compared women prescribed or issued NRT with those who were not and were derived from five discrete birth cohorts. A UK cohort reported outcomes in two manuscripts [32,34] and a PhD [38]; a Danish cohort reported outcomes in five papers [33,35,37,39,42] and Canadian [40], US [36] and Australian [41] cohorts were reported in single studies. Eleven studies described

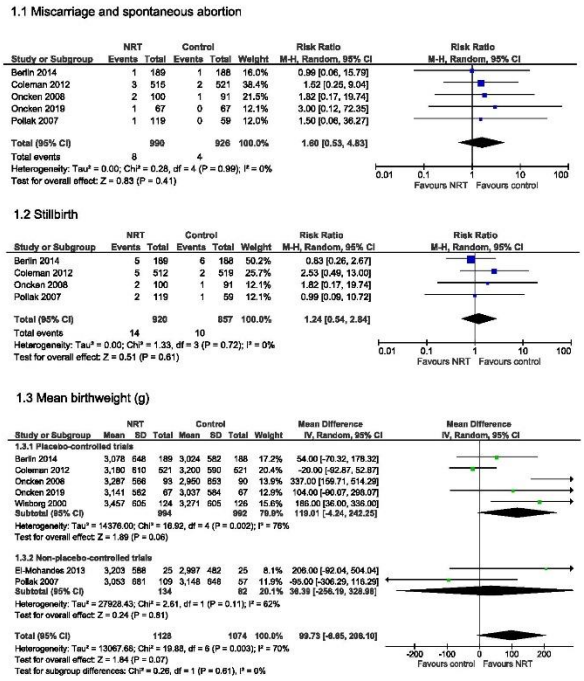

Figure 1 Meta-analyses of randomized controlled trials (RCTs) (from Cochrane Review). [Colour figure can be viewed at wileyonlinelibrary.com]

NRT administration to small, experimental interventional cohorts of inpatient pregnant women who usually smoked but were temporarily abstinent [43–53], and were based in Sweden [43,48], the United States [45,46,49–53], the United Kingdom [44] and Finland [47]. These mainly compared short-term fetal and maternal physiological observations when abstinent and using NRT to those when women smoked. The final Danish study was interventional; participants were a subgroup of women in a quasi-RCT who had been offered and accepted NRT [54].

Maternal age was reported by 17 studies [32–36,38–40,43,45,46,48–53] and used as a confounder in analyses, but not reported in three [33,37,41,42]. Socio-economic status or education level was reported by 11 studies [32–40,42,51]. Maternal comorbidities were included as confounders in six routine health-care studies [32,34,37,38,40,41], and as exclusion criteria in five interventional studies [46,50–53]. One interventional study reported comorbidities for each participant and analysed data by condition [44].

NRT exposure data was obtained from electronic medical records or prospectively from telephone interviews in nine routine health-care studies [32–35,37–39,41,42]; two others collected data retrospectively via self-administered postal questionnaires [36,40] sent 3–8 years [40] and 2–3 months [36] after pregnancy. Although women in the Danish cohort were asked in which gestational weeks they had used NRT or smoked, manuscripts did not report the details [33,35,37,39,42] and one routine health-care study reported median duration of NRT use but not when, in pregnancy, this occurred [40]. All 12 interventional studies reported women's gestational ages at NRT administration, with nine providing mean gestational ages at exposure (range = 21.5–35.6 weeks) [45,46,48–54].

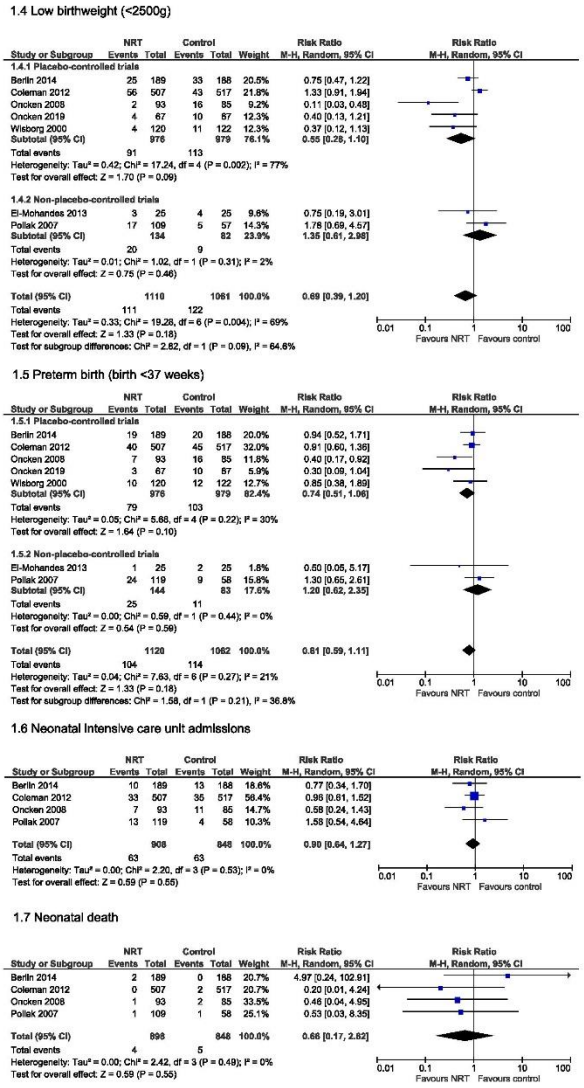

Figure 1 Continued.

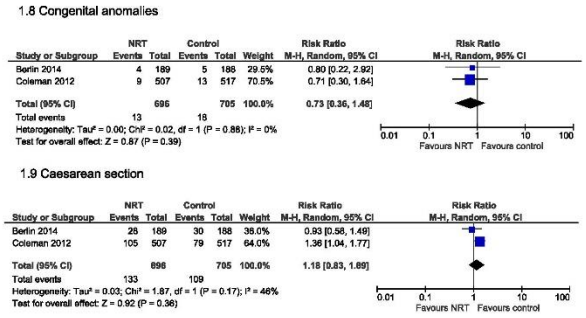

Figure 1 Continued.

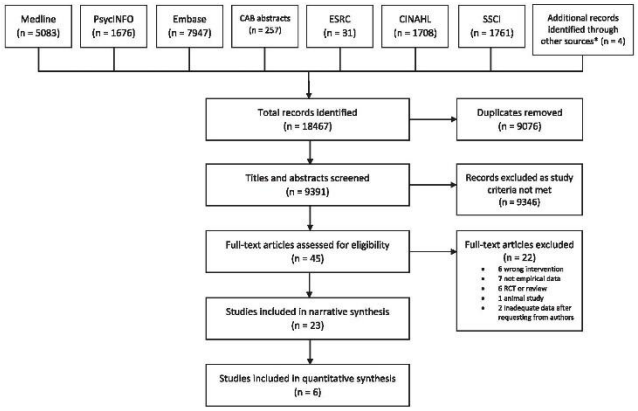

Figure 2 Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) flow diagram: non-RCT review. PRISMA [17] flow diagram showing study selection and reasons for study exclusion. *Hand searches of references located, a PhD thesis and an unpublished study at the time of searching known to the authors.

Table 2 shows which studies reported or adjusted for women's smoking before or after NRT use. Of the six studies in meta-analyses, three reported women's smoking behaviour before NRT use/exposure [33,36,39], but data were collected by questionnaire at set time-points, so no smoking behaviour information was available later in participants' pregnancies. Consequently, many pregnant women in NRT-exposed arms of meta-analyses will also have smoked, and exposures to NRT and smoking are not completely differentiated. Two routine health-care studies adjusted for smoking status during NRT use [33,37]. Two routine health-care studies recruited only pregnant women who smoked, and investigated impacts of using NRT within this group [40,41]. Experimental studies all recorded women's smoking status at the time of recruitment, and nine also validated abstinence just before NRT was given to women [43–46,48–52] and two followed participants until childbirth, collecting some information on smoking after NRT exposure [53,54].

Table 3 summarizes studies' outcomes. Routine health-care cohorts reported pregnancy outcomes such as congenital anomalies [34,35], birth weight [36–40], gestational age at birth [36,37,39,40] and stillbirth [32,33]. Interventional studies generally monitored physiological observations, including biophysical profiles [49,52], umbilical and uterine artery Dopplers

Table 2 Characteristics of included non-RCTs.

First author, Location Date	*Recruitment method (location/setting, eligibility criteria, data collection)*	*Study objectives*	*Baseline characteristics/group differences*
Cohort studies of NRT use during routine health-care – with control/unexposed group for analysis (non-smokers who did not use NRT)			
Outcome: Stillbirth			
Dhalwani UK 2018 [32]	The Health Improvement Network (THIN) database UK database of anonymized electronic primary care records Longitudinally prospectively collected data from 570 GP surgeries, representing 6% of UK population Singleton pregnancies with deliveries between 2000 and 2013	To compare risk of stillbirth between pregnant smokers, non-smokers and those prescribed NRT	220630 singleton pregnancies ~50% in NRT and smoking groups in two most deprived quintiles versus ~25% of controls. Women aged 15–49 years. Higher rates of mental illness in NRT and smoking groups
Strandberg-Larsen Denmark 2008 [33]	Danish National Birth Cohort (DNBC) Population-based cohort of pregnant women and their offspring, recruitment 1996–2002 50% of Danish GPs participated and invited pregnant women to participate. An estimated 60% of invited women took part Eligible if: pregnant, Danish speaker, planned to carry pregnancy to term Computer-assisted telephone interview between 12–16/40 where women self-reported NRT use and smoking behaviours during pregnancy Outcomes identified from Civil Registration System, Danish Medical Birth Registry using record linkage, National Hospital discharge register or by self-report	To examine if NRT use in pregnancy has an association with the risk of stillbirth	Pregnancies 100418 Information from first interview 90165 After exclusion: 87032 singleton pregnancies mean gestational age at: recruitment 11.5 weeks; interview 16.9 weeks. Women using NRT more often at least 35 years, nulliparous, BMI $< 25\ kg/m^2$, alcohol consumption ≥ 2 drinks/week, higher caffeine intake
Outcome: major congenital anomalies (MCAs)			
Dhalwani UK 2015 [34]	THIN database (see Dhalwani 2018) Pregnancy cohort created from all 15–49-year old women in database by linking pregnancy and birth-related codes in women's medical records to live births of children from January 2001 to December 2012 Major congenital anomaly (MCA) diagnoses extracted from European Surveillance of Congenital Anomalies and Twins classification system (EUROCAT). Minor CAs and MCAs specifically attributed to known teratogens excluded	To assess the relationship between early pregnancy exposure to NRT or smoking with MCA in offspring	192498 live-born children. Mothers in smoking group younger compared to NRT group/controls. ~25% mothers in NRT/smoking groups in most deprived quintile. NRT/smoking groups had higher proportions of maternal morbidities (asthma, mental illness)

(*Continues*)

Table 2. (Continued)

First author, Location, Date	*Recruitment method (location/setting, eligibility criteria, data collection)*	*Study objectives*	*Baseline characteristics/group differences*
Morales-Suárez-Varela Denmark 2006 [35]	DNBC (see Strandberg-Larsen, 2008). Based on first child in the cohort, women included from January 1997 to December 2003 Included: those who answered first interview smoking questions (11–25 weeks) Excluded: women with ovarian/cervical cancer during pregnancy, multiple pregnancy Congenital malformation data obtained from the Medical Birth Registry and hereditary diseases and chromosome anomalies diagnosed during first year of life obtained from the National Hospital Discharge Registry. Congenital malformations classified as per EUROCAT criteria	To examine whether maternal smoking and use of nicotine substitutes during the first 12 weeks of pregnancy increased the prevalence of congenital malformations	76 768 live-birth singleton pregnancies included Smokers younger, lower body weight, higher alcohol intake and less well educated than non-smokers. NRT members of cohort not described separately
Outcomes: gestational age at birth (including preterm birth), birth weight (including mean and low birth weight)			
Gaither USA 2009 [36]	2004 Phase V Pregnancy Risk Assessment Monitoring System (PRAMS)—population-based surveillance system with information on maternal behaviour before, during and after pregnancy Participating states draw stratified sample of new mothers from birth certificates and send self-administered questionnaires on exposures in pregnancy. Birth outcome information from birth certificates 4 states included in the study—Colorado, Louisiana, Maine, Washington. Excluded if: no prenatal care, missing data on marital status, education, parity, Medicaid, alcohol use	To examine factors associated with the prescription or advisement of NRT during pregnancy and its association with adverse pregnancy outcomes	6041 women aged 18–45 completed questionnaires After exclusion criteria: 5716 women ≤ 24 years less likely to be prescribed/recommended NRT compared to women ≥ 35 years. Women who gained ≤ 15 lbs significantly more likely to be prescribed NRT
Lassen Denmark 2010 [37]	DNBC (see Strandberg-Larsen 2008) Interviews at, on average, 16/40 and 31/40 Included if gave birth to live-born singletons. Excluded if: gave birth before 28 completed weeks; uncertain gestational age/estimated date of delivery; interview before 27 completed weeks; information missing from the variables included in multivariate analysis, did not answer both interviews Information for birth outcomes obtained by linking of cohort to National Patient Registry	To estimate the association between the use of NRT during pregnancy and offspring birth weight	101 042 pregnancies in cohort After exclusion criteria applied: 72 761 Median tobacco use in NRT users was 9.5 pack weeks, 88% of NRT users smoked during 1 or more weeks in first 2 trimesters NRT users had lower socio-economic status than non-smokers but higher than smokers

(*Continues*)

Table 2. (Continued)

First author, Location Date	*Recruitment method (location/setting, eligibility criteria, data collection)*	*Study objectives*	*Baseline characteristics/group differences*
Outcomes: low and mean birth weight, fetal death, mode of delivery			
Dhalwani UK 2014 [38]	Chapter 7 of PhD thesis THIN database (see Dhalwani 2018) Pregnancy cohort created from all 15–49-year-old women in database by linking pregnancy and birth-related codes in women's medical records Low and mean birth weight measured from live births January 2001–December 2009. Stillbirth or fetal death from January 2001–September 2009	To examine the association between NRT and smoking exposure and risk of fetal death, birth weight and mode of delivery	*n* variable for populations depending on outcome—see Results column Approximately 50% mothers in both smoking and NRT groups in 2 most deprived quintiles
Outcome: infantile colic			
Milidou Denmark 2012 [39]	DNBC (see Strandberg-Larsen 2008) 2 computer-assisted telephone interviews in 2nd and 3rd trimesters and at 6 months postpartum concerning infant's behaviour, development, nutrition and frequency and duration of cry episodes Included if completed both 2nd trimester and postnatal interviews. Based on first child in the cohort, so 3695 younger siblings excluded Exposure to nicotine was self-reported smoking or use of NRT during pregnancy	To compare the association between intrauterine exposure to tobacco smoke and infantile colic with the possible association between NRT and infantile colic	101 042 pregnancies After exclusion: 63 128 mother–infant dyads included Mothers that smoked either with/without NRT tended to be less well educated and slightly younger in age
Cohort studies of routine health care—with no control group for analysis (non-smokers who did not use NRT)			
Outcomes: preterm birth and small for gestational age			
Bérard Canada 2016 [40]	Quebec Pregnancy Cohort (Canada). All pregnancies between January 1998 and December 2009 recorded 1288 women who reported smoking just prior to pregnancy from 6732 of 8505 women selected randomly and contacted annually to fill in a standardized questionnaire (sent in the post, twice, 3– 8 years after pregnancy of interest) on smoking status and information on NRT use NRT usage obtained from Régie de l'Assurance Maladie du Québec medication database. Cross-checked with NRT and smoking status from questionnaire	To quantify the effect of gestational use of bupropion[4] and nicotine patch replacement therapy on the risk of prematurity and small for gestational age	6732 women completed the questionnaire 1288 women were smokers before pregnancy and therefore eligible for inclusion Women aged 15–45 years, NRT initiated in first trimester

(*Continues*)

Table 2. (Continued)

First author, Location Date	*Recruitment method (location/setting, eligibility criteria, data collection)*	*Study objectives*	*Baseline characteristics/group differences*
Tran Australia 2020 [41]	Smoking MUMS study Population-based cohort focused on pharmacotherapies for smoking cessation All pregnancies resulting in a birth (> 20 weeks) across two states, New South Wales and Western Australia, 2004–12 Linked records from 4 data sources—perinatal data, dispensing data for pharmaceuticals subsidized through benefits scheme, hospital admission and deaths Smoking status from perinatal or maternal hospital admission data	To compare risk of adverse perinatal outcomes between pregnancies exposed to pharmacotherapies (NRT, bupropion, varenicline[a]) and pregnancies exposed to smoking but no pharmacotherapy	3608 pregnancies exposed to either NRT or smoking Well-balanced baseline characteristics between groups as pregnancies matched between NRT users and smokers
Outcome: strabismus in infants			
Torp-Pedersen Denmark 2010 [42]	DNBC (see Strandberg-Larsen 2008). Children born between 1996–2003. Interviews ×2 during pregnancy, 6 and 18 months postpartum to obtain information on exposures of potential relevance National Patient Register and Health Security System identified children with strabismus diagnosis, previous strabismus surgery or evaluation for strabismus. Medical records for cases evaluated by 2 experienced orthoptists then linked with information from birth cohort	To investigate the effect of *in-utero* exposure to maternal smoking and consumption of alcohol, coffee and tea on the risk of strabismus	Review of 5635 medical charts yielded 1320 instances of strabismus

(*Continues*)

Table 2. (Continued)

First author, Location Date	*Exposure group(s) (including dose and route of NRT)*	*Non-exposed group*	*Outcome Measures (i.e. adverse fetal/ maternal health outcomes)*	*Main/significant results*
Cohort studies of NRT use during routine health-care – with control/unexposed group for analysis (non-smokers who did not use NRT)				
Outcome: Stillbirth				
Dhalwani UK 2018 [32]	NRT[a], n = 5221, women prescribed NRT during pregnancy or before conception as recorded in primary care record. Average NRT prescription duration 2 weeks (IQR: 6 days–2 weeks), 80% within first 2 trimesters. No quantification of smoking before/during/after NRT exposure. Smokers n = 18 407, smoking recorded in medical record with no quantification of exposure	Controls n = 197 002, recorded as non-smoker, or not smoked for 3 years in medical records	Stillbirth—baby born with no signs of life at or after 28 weeks gestation	Stillbirth, n = 805 (3.6 in 1000 births) NRT n = 26 (0.5%), smokers n = 96 (0.52%), controls n = 683 (0.35%) When compared with controls: NRT OR = 1.44 (CI = 0.97–2.14), Smokers OR = 1.52 (CI= 1.23–1.89), $P < 0.01$ Similar findings after adjustment for confounders (age, socio-economic status, BMI, diabetes) Higher prevalence of stillbirth if maternal age > 35 years or diabetes
Strandberg-Larsen Denmark 2008 [33]	NRT user[b] n = 1927, any use of NRT between LMP and interview Subcategorized into: NRT user + smoker n = 1091, where participants said they had co-exposure to NRT and smoking at time of interview, by grams of tobacco on average/ day or week NRT user + non-smoker n = 836, non-smokers and those who smoked in pregnancy but were ex-smokers at time of interview and used NRT Smokers n = 13 266, those who reported smoking during pregnancy at time of interview (group including ≤ 10 g/ tobacco/day and > 10 g/day) Unclear about smoking behaviour after interview (apart from 836 who were not smoking afterwards)	Non-users of NRT n = 85 105 Sub-categorized into: smokers (see previous column) and non-smokers (controls) n = 71 839, non-smokers who were not exposed to NRT, including those who quit before conception, or reported being an ex-smoker at interview but may have smoked in early pregnancy	Stillbirth—any fetus that did not breathe or show any other sign of life at birth > 20 weeks gestation	Stillbirth n = 495 (rate 5.7 in 1000 births) Compared to non-users of NRT, overall NRT user stillbirth hazard ratio (HR) = 0.57 (CI = 0.28–1.16); adjusted for maternal age, household socio-occupational status and smoking habits) Subcategories adjusted for maternal age and socio-occupational status: controls, n = 380 (reference for all HRs) NRT user + smoker—HR = 0.83 (CI = 0.34–2.00) NRT user + non-smoker, HR = 0.67 (CI = 0.21–2.08) Smokers—HR = 1.46 (CI = 1.17–1.82). Definition of stillbirth changed to no signs of life at > 22 weeks—no significant change in results

(*Continues*)

Table 2. (Continued)

First author, Location, Date	*Exposure group(s) (including dose and route of NRT)*	*Non-exposed group*	*Outcome Measures (i.e. adverse fetal/ maternal health outcomes)*	*Main/significant results*
Outcome: major congenital anomalies (MCAs)				
Dhalwani UK 2015 [34]	NRT[a] n = 2677, women prescribed NRT either during first trimester of pregnancy or within 4 weeks before estimated conception date according to primary care record. No quantification of smoking before/during/after NRT exposure Smokers n = 9980, smoking recorded in medical record with no quantification of exposure	Controls n = 179 841, recorded as non-smoker, or not smoked for 3 years in medical records	MCAs: heart; limb; genital system; urinary system; chromosomal; musculoskeletal; orofacial cleft; digestive system; nervous system; other malformations; eye; respiratory system; genetic; abdominal wall; ear, face and neck	Maternal age at conception, socio-economic status and preconception BMI similar in women with infants with MCAs and those without. MCA group had a slightly higher proportion of maternal morbidities. At least 1 MCA n = 5535 (288 per 10 000 live births) All MCAs combined: NRT n = 90 (336 of 10 000), smokers n = 314 (315/10 000), controls n = 5131 (285 of 10 000) when compared with controls across all MCAs (adjusted for maternal age at conception, Townsend deprivation index score, maternal diabetes, asthma, mental illness and multiple births) NRT OR = 1.12 (CI = 0.84–1.48), smokers OR = 1.05 (CI = 0.89–1.23) There were no statistically significant associations between maternal NRT use and system-specific anomalies except for respiratory anomalies OR (compared with controls) = 4.65 (99% CI = 1.76–12.25, $P < 0.001$)
Morales-Suárez-Varela Denmark 2006 [35]	NRT[b] n = 250, women who reported during interview using nicotine substitutes in the first 12 weeks of pregnancy but did not smoke. No information available on smoking behaviour after NRT exposure as questionnaire only asked about use in first 12 weeks of pregnancy. Smokers n = 20 603, women who reported smoking during the first 12 weeks of pregnancy at interview; note n = 3791 missing data unaccounted for, so for analysis n = 16 812 Quantified as ≤ 10/day and > 10/day.	Controls n = 55 915 women who reported being non-smokers who did not use NRT	Various congenital malformations: nervous system: ear, eye, face and neck; circulatory system; respiratory system; cleft lip and palate; digestive system; genital organs; urinary system; musculoskeletal system; chromosomal anomalies; other	Children born with all congenital malformations n = 3767 For major malformations: controls n = 2186 [reference for all relative prevalence rate ratios (RPR)] NRT n = 11 RPR = 1.13 (CI = 0.62–2.07) Smokers n = 722 < 10/day n = 535, RPR = 1.12 (CI = 1.02–1.23) > 10/day n = 187, RPR 1.09 (CI = 0.93–1.27) NRT and major musculoskeletal RPR = 2.05 (CI = 0.91–4.63)

(*Continues*)

Table 2. (Continued)

First author, Location Date	*Exposure group(s) (including dose and route of NRT)*	*Non-exposed group*	*Outcome Measures (i.e. adverse fetal/ maternal health outcomes)*	*Main/significant results*
Outcomes: gestational age at birth (including preterm birth), birth weight (including mean and low birth weight)				
Gaither USA 2009 [36]	NRT[a] n = 225, smoking participants self-reported if a health-care professional prescribed a nicotine spray, inhaler or pill, or recommended using a nicotine patch or gum during prenatal visits as a means of smoking cessation Smokers n = 637 women who reported smoking during pregnancy but no NRT use No quantification of smoking or smoking intensity in either group, before or after NRT exposure.	Non-smokers n = 4854 women who reported not smoking during pregnancy and therefore were not recommended or prescribed NRT	Preterm birth—birth at <37 weeks gestation Low birth weight (LBW) - < 2500 g at birth	Low birth weight: NRT n = 84 (13.05%). Smokers n = 205 (9.26%), Non-smokers n = 1303 (6.99%) Unadjusted OR when compared with non-smokers: NRT OR = 2.00 (CI = 1.13–3.45), smokers OR = 1.36 (CI = 0.98–1.88). Similar findings after ORs adjusted for age, marital status, education and ethnicity Preterm birth: NRT n = 66 (17.54%), smokers n = 156 (10.19%), non-smokers n = 1165 (9.42%). Unadjusted ORs when compared with non-smokers: NRT OR = 2.05 (CI = 1.16–3.60), smokers OR = 1.09 (CI = 0.76–1.57.) Similar findings after ORs adjusted for age, marital status, education and ethnicity
Lassen Denmark 2010 [37]	NRT[a] n = 1828, women asked if they had used NRT, which type and in which weeks of pregnancy (First 27 weeks was exposure period of interest). Smoking quantified in pack weeks. For analysis, smoking information required, so n without missing smoking info, NRT n = 1753 Smokers n = 15796, women who reported smoking but no NRT use, by pack weeks. Smoking not quantified again after questionnaire completed	Non-smokers n = 53771, self-report of being a non-smoker not using NRT	Gestational age at birth split into: Preterm <259 days Term 259–293 days Post-term >293 days Birth weight (change in mean birth weight in g)	Gestational age at birth: Preterm: NRT 4.1%, smokers 3.9%, non-smokers 3.2%. Term: NRT 87.2%, smokers 87.0%, non-smokers 87.7%. Post-term: NRT 8.8%, smokers 9.1%, non-smokers 9.1% Change in mean birth weight in g after all NRT use within the first 27 completed weeks of gestation 0.25 (CI = 2.31–2.81), adjusted for gestational age, smoking status, partner smoking status, parity, pre-pregnancy BMI, height, alcohol, coffee, exercise, infant sex, socio-economic status, weight loss, eating disorder, fertility problems, vaginal bleeding, nausea, hypertension Subgroups change in mean birth weight in g. Nicotine patch −4.37 (CI = −13.34–4.60). Nicotine gum 0.48 (CI = − 2.51–3.48). Nicotine inhaler 6.19 (CI = − 0.40–12.79) More than one product −10.72 (CI = −26.51–5.05)

(*Continues*)

Table 2. (Continued)

First author, Location, Date	*Exposure group(s) (including dose and route of NRT)*	*Non-exposed group*	*Outcome Measures (i.e. adverse fetal/ maternal health outcomes)*	*Main/significant results*
Outcomes: low and mean birth weight, fetal death, mode of delivery				
Dhalwani UK 2014 [38]	NRT[a], women prescribed NRT during pregnancy or before conception according to primary care record. No quantification of smoking before/during/after NRT exposure Smokers, smoking recorded in medical record with no quantification of exposure	Controls, recorded as non-smoker, or not smoked for 3 years in medical records	Low birth weight (<2500 g). Fetal death—combined stillbirth and miscarriage data Mode of delivery. Change in mean birth weight	For each outcome: *N* = population, *n* = outcome, adjusted ORs (99% CI), *P*-value. low birth weight [total *N* = 96782 (*n* = 7739 unknown exposure so excluded)]: NRT *N* = 1223, *n* = 135, OR = 1.88 (1.42–2.49), $P < 0.001$ Smokers *N* = 4622, *n* = 435, OR = 1.73 (1.45–2.06), $P < 0.001$. Non-smoker *N* = 13088, *n* = 740, reference for ORs fetal death [total *N* = 111802 (*n* = 10575 unknown exposure so excluded)]: NRT *N* = 5234, *n* = 491, OR = 0.44 (0.38–0.50), $P < 0.001$ Smokers *N* = 50643, *n* = 10560, OR = 1.16 (1.11–1.21), $P < 0.001$ Non-smoker *N* = 150175 (*n* = 25962, reference for ORs) Mode of delivery (adjusted RRR relative to normal delivery/non-smokers) NRT Assisted RRR = 0.68 (0.54–0.85), $P < 0.001$; C-section RRR = 0.92, (0.81–1.05), $P = 0.120$ Smokers assisted RRR = 0.76 (0.68–0.86), $P < 0.001$; C-section RRR = 0.88 (0.81–0.95), $P < 0.001$ Mean (SD) birth weight 3.41 kg (0.59). Gestational age for same pregnancies 40 weeks (2.11). Change in mean birth weight with NRT exposure compared with non-smokers β = −168 g (−214, −122), $P < 0.001$. Change in mean birth weight after smoking compared with non-smokers β = −139 g, (−165, −113), $P < 0.001$

(*Continues*)

Table 2. (Continued)

First author, Location Date	*Exposure group(s) (including dose and route of NRT)*	*Non-exposed group*	*Outcome Measures (i.e. adverse fetal/ maternal health outcomes)*	*Main/significant results*
Outcome: infantile colic				
Mildou Denmark 2012 [39]	NRT users[b] n = 207—those who reported NRT use with no smoking during pregnancy N = 194 used one of: gum, patch or inhalator. N = 13 used a combination of NRT types Smokers using NRT (in combination) n = 1245 Smokers n = 15016, those who reported smoking without using NRT Smoking quantified as cigarettes/day, data not given per exposure group. Smoking behaviour elicited at postnatal interview but data not given	Controls n = 46660, unexposed i.e. no smoking and no NRT	Infantile colic: Wessel's criteria used—crying or fussing for more than 3 hours a day for more than 3 days a week, starting before age 3 months Preterm birth (gestational age at birth < 37 weeks) and low birth weight (birth weight < 2500 g) also reported for all groups	Preterm birth: controls 4.0%, NRT users 2.9%, smokers 4.9%, smoking and NRT 5.2% Low birth weight: controls 2.4%, NRT users 2.9%, smokers 4.3%, smoking and NRT 4.8% Infantile colic—n = 4974 (7.9%) of total infants. Crude unadjusted ORs when compared with controls n = 3397: NRT n = 23, OR = 1.6; smokers[c] n = 1417, OR = 1.3; smoking and NRT n = 137, OR 1.6 Similar results (all significant) when adjusted for maternal age, first parity, daily coffee consumption, weekly alcohol consumption, and binge-drinking episodes and when further adjusted for couple's combined educational and occupational status
Cohort studies of routine health care—with no control group for analysis (non-smokers who did not use NRT)				
Outcomes: preterm birth and small for gestational age				
Bérard Canada 2016 [40]	NRT n = 316, smokers with prescription for NRT patch filled before 1st day of gestation with duration overlapping beginning of pregnancy, or reporting OTC NRT patch use Some quit smoking during NRT use but no quantification of this in paper Median duration of use 54 days (IQR: 36–72 days) Smokers n = 900, pregnant smokers without bupropion or NRT patch exposures during pregnancy. No quantification of smoking	No controls—there was no group exposed to neither NRT nor smoking	Prematurity— birth before the 37th week of gestation Small for gestational age (SGA): lowest 10th percentile of the gestational age-specific birth weight in the cohort	Preterm birth: NRT n = 25 (7.9%), smokers n = 240 (26.7%) SGA: NRT n = 44 (13.9%), smokers n = 149 (16.6%) ORs for NRT compared with smoking—adjusted for maternal socio-economic status, health care utilization and comorbidities before pregnancy Preterm birth: NRT OR = 0.21 (CI = 0.13–0.34); SGA: NRT = 0.61 (CI = 0.41–0.90) Mean gestational age in weeks (SD): NRT = 38.9 (1.9), smokers 37.5 (3.3). Mean birth weight in g (SD): NRT = 3257.9 (553.1), smokers 2943.5 (733.5) Smokers had significantly shorter gestation and lower birth weight newborns compared with bupropion or NRT users ($P < 0.05$)

(*Continues*)

Table 2. (Continued)

First author, Location Date	Exposure group(s) (including dose and route of NRT)	Non-exposed group	Outcome Measures (i.e. adverse fetal/ maternal health outcomes)	Main/significant results
Tran Australia 2020 [41]	NRT n = 328, one or more dispensings of NRT in 100 days prior to conception or in gestation period. Gestational age at exposure calculated. All women assumed to be smokers but degree of smoking alongside NRT use not quantified. Smokers n = 3280 (10 : 1 matching), smoking recorded in perinatal data collection ('yes' or smoking > 0 cigarettes in 1st or 2nd halves of pregnancy) or if hospital admission data stated smoked in the last month before delivery	No controls—there was no group exposed to neither NRT nor smoking	Preterm birth, SGA, Apgar at 5 min < 7, admission to neonatal special care, severe neonatal morbidity complications, emergency caesarean, severe maternal morbidity complications, PPROM, placental abruption, perinatal death (stillbirth or neonatal death)	Any adverse perinatal event (composite of all 10 events): NRT n = 147, smokers n = 1520, HR = 1.02 (CI = 0.84–1.23). Preterm birth: NRT n = 36, Smokers n = 358, HR = 1.00 (CI = 0.71–1.42) SGA: NRT n = 47, Smokers n = 578, HR = 0.77 (0.56–1.07) Admission to neonatal special care: NRT n = 66, Smokers n = 692, HR = 0.97 (CI = 0.74–1.26) Severe neonatal morbidity: NRT n = 27, Smokers n = 345, HR = 0.82 (CI = 0.55–1.23) Emergency caesarean: NRT n = 37, Smokers n = 421, HR = 1.01 (CI = 0.70–1.45). Severe maternal morbidity complications: NRT n = 10, Smokers n = 79, HR = 1.22 (CI = 0.60–2.46). PPROM: NRT n = 16, Smokers n = 139, HR = 1.15 (CI = 0.68–1.94). Apgar 5 min < 7: NRT n = 6, smokers n = 105, HR = 0.59 (CI = 0.25–1.37). Placental abruption: NRT n < 5,smokers n = 34. Perinatal death: NRT n < 5, Smokers n = 30
Outcome: strabismus in infants				
Torp-Pedersen Denmark 2010 [42]	NRT n = 61, NRT use self-reported at least once during pregnancy via maternal interview, smoking status unknown Smokers n = 415. Not directly compared with NRT, quantified by average cigarettes/ day during pregnancy	Unexposed to NRT n = 1239, women who did not use NRT (although smoking status not reported)	Strabismus in infants	The use of nicotine replacement therapy in pregnancy was associated with a non-significant 22% increase (RR = 1.22, CI = 0.92–1.61) in strabismus risk in comparison with no maternal NRT use Smoking was associated with a significantly increased risk compared with mothers who did not smoke (RR = 1.26, CI = 1.11–1.43). Higher numbers of cigarettes/day associated with higher risk of strabismus

(*Continues*)

Table 2. (Continued)

First author, Location, Date	*Study design*	*Recruitment method (location/setting, eligibility criteria, data gathering)*	*Study objectives*	*Baseline characteristics/group differences*	*Intervention group(s) (including dose and route of NRT)*
Cohorts with NRT administered as an intervention in experimental settings					
Outcome: Fetal breathing movements during nicotine administration					
Gennser Sweden 1975 [43]	Controlled longitudinal cohort study	Department of Obstetrics and Gynecology, University of Lund Performed in morning hours after breakfast following a smoking-free interval of at least 12 hours. Women in recumbent position, with equilibration period of 15 minutes of fetal monitoring before a control period of 30 minutes. Then either tobacco/non-tobacco cigarette or gum smoked over 5 minutes, respectively. Fetal breathing movements measured from start to 60 minutes after intervention using ultrasonographic technique, along with maternal breathing and ECG	To study the influence of maternal smoking on the fetal breathing movements and attempts to determine the role of nicotine in this response	12 healthy pregnant women aged 20–31 years between 33rd–39th gestational weeks All smokers, average cigarette consumption 7–20 cigarettes a day	$N = 6$ tested with: ● Standard cigarette (1.7–1.8 mg nicotine) ● Nicotine chewing gum 2 mg ● Nicotine chewing gum 4 mg One substance tested per day on 3 consecutive days
Manning UK 1976 [44]	Uncontrolled longitudinal cohort study	Nuffield Institute for Medical Research, University of Oxford Observations made in morning after patient had eaten normal breakfast and had abstained from smoking overnight. Fetal breathing movements recorded using A-scan ultrasound method for at least 30 minutes before smoking/chewing and for at least 60 minutes afterwards	To determine the factor in cigarette smoke and other substances responsible for the depression of fetal breathing	64 women, all in third trimester, all chronic smokers	Nicotine gum $n = 12$ (chewed vigorously for at least 20 minutes): ● 1×4mg nicotine containing gum ($n = 7$) ● 2×4mg nicotine containing gum ($n = 5$)
Outcome: fetal blood flow during nicotine administration					
Bruner USA 1991 [45]	Prospective cross-over cohort study	Obstetric clinics of the Hospital of the University of Pennsylvania. Large, indigent, inner city	Prospective comparison of effects of maternal smoking of a single cigarette and buccal nicotine	47 healthy pregnant women ● Non-smokers ($n = 16$) ● Prior smokers ($n = 8$)	NRT: phase 2: 1–2 weeks after phase 1, same participants chewed one piece of nicotine

(*Continues*)

Table 2. (Continued)

Cohorts with NRT administered as an intervention in experimental settings						
First author, Location, Date	*Study design*	*Recruitment method (location/setting, eligibility criteria, data gathering)*	*Study objectives*	*Baseline characteristics/group differences*	*Intervention group(s) (including dose and route of NRT)*
		population of women in late 2nd and 3rd trimesters of pregnancy from menstruation dates and confirmation ultrasound Inclusion criteria specified. Randomly approached, offered written, informed consent. All examinations 8–11 a.m., after abstaining from smoking for 30 minutes. Resting in recumbent position until HR/BP at constant baseline. Uterine/umbilical artery FVWs measured at baseline and after intervention (at 1 minute and 10 minutes) using continuous wave Doppler transducer	exposure on uterine and umbilical artery blood FVWs	● Smokers ($n = 23$, ≤ 10 cigarettes/day $n = 13$, 11–20/day $n = 10$) Mean maternal age 24 years. Mean gestational age at study entry 30/40, 33/40 at gum phase	polacrilex (gum) containing 2 mg nicotine vigorously 15 times and then tucked into cheek until FVWs recorded
Onchen USA 1997 [46]	Prospective cross-over cohort study	Recruitment: April–October 1995 through newspaper advertisements and local physician practices Inclusion and exclusion criteria specified Expired CO measured before patch placement/smoking to verify abstinence, and at 2, 3, 4, 5, 6 and 8 hours. Maternal BP, HR and ultrasound resistance indices (RIs) determined at 2, 3, 4, 6 and 8 hours after patch applied/smoking began. 20-minute FHR monitoring performed at baseline and repeated 4 hours later	To compare nicotine concentrations and fetal middle cerebral artery RIs during transdermal NRT with smoking	23 evaluated for suitability to take part 17 subjects met eligibility criteria 2 exclusions: one because of a single umbilical artery; the other quit smoking after the first session Mean gestational age 28 + 3/40 (SD ±20 days)	Nicotine patch $n = 15$, 21 mg patch applied for 8 hours after abstaining from 8 p.m. the night before the study Women randomized to one intervention and 1 week later crossed over to other study arm
Outcomes: maternal and fetal physiological observations during NRT administration					
Lehtovirta Finland 1983 [47]	Before–after study – two separate cohorts for two interventions	Department of Obstetrics and Gynecology, University Central Hospital, Helsinki, Finland	To determine the effects of nicotine and nicotine-free herbal cigarettes	31 healthy pregnant women: 15 2nd trimester 24–26 weeks 16 3rd trimester 37–40 weeks	NRT $n = 15$ chewed nicotine gum containing 2 mg nicotine

(*Continues*)

Table 2. (Continued)

Cohorts with NRT administered as an intervention in experimental settings

First author, Location, Date	*Study design*	*Recruitment method (location/setting, eligibility criteria, data gathering)*	*Study objectives*	*Baseline characteristics/group differences*	*Intervention group(s) (including dose and route of NRT)*
		Subjects rested in 15 degrees left lateral recumbent position. Abdominal fetal electrocardiography recorded by a cardiotocograph. FHR analysis performed after 5-minute period to establish normal variability and measured for 1 minute at 15 minutes before, during and 25 minutes after administration of nicotine chewing gum or smoking a herbal cigarette Baseline FHR estimated visually from cardiogram. Maternal BP/HR also measured for 1-minute periods.	on quantitated FHR variability during the antepartum period	8 regular smokers, 23 had stopped smoking before or at the start of pregnancy	for 20 minutes (7 in 2nd trimester, 8 in 3rd)
Lindblad Sweden 1987 [48]	Prospective cross-over cohort study	Recruitment methods not documented. Uncomplicated pregnancies, singleton fetus in cephalic position Women abstained from smoking for at least 12 hours before study. During study, women in left lateral position. Maternal HR, BP, FHR and fetal blood velocity were measured; 3 control measurements taken and then gum chewing started. Recordings were made every 5 minutes for 45 minutes after chewing started	To evaluate maternal and fetal haemodynamics after exposure to nicotine	20 pregnant smokers Mean consumption 12 cigarettes/day (SD = 5.3) Mean maternal age 30.1 years (SD = 3.8), mean gestational age at time of study 35.6 weeks (SD = 2.2), 7 women primiparous	NRT $n = 20$, chewed nicotine gum containing 4 mg nicotine for 30 minutes Women randomized to one intervention and one day later crossed over to other study arm
Ogburn Jr USA 1999 [49]	Uncontrolled before–after cohort study	Pregnant cigarette smokers in 3rd trimester recruited from Department of Obstetrics at the Mayo Clinic	To determine in abstinent pregnant smokers whether nicotine patch therapy acutely compromised fetal wellbeing	23 women enrolled 2 subjects discontinued before inpatient phase Means at baseline outpatient visit (±SD):	NRT $n = 21$, 22 mg/24-hour nicotine patch was applied each morning (replaced daily)

(*Continues*)

Table 2. (Continued)

Cohorts with NRT administered as an intervention in experimental settings

First author, Location, Date	*Study design*	*Recruitment method (location/setting, eligibility criteria, data gathering)*	*Study objectives*	*Baseline characteristics/group differences*	*Intervention group(s) (including dose and route of NRT)*
		Inclusion and exclusion criteria specified. Data collection: screening questionnaires/consent. Then bloods, urinalysis, ECG, expired CO levels and physician examination, ultrasound for gestational age, fetus size and umbilical artery Doppler studies. On another day, baseline expired CO/bloods, FHR, umbilical artery Doppler and biophysical profile taken before 1st cigarette. Returned after smoking *ad libitum* all day for repeat tests. Within 7 days, admitted in evening after day of smoking, for 4 days/nights to smoke-free setting. Each morning, further tests (as above) and then repeated after 8 hours of NRT patch use		● Age 26.5 years (±5.7) ● Gestational age 27.4 weeks (±2.7) ● Cigarettes/day at time of study 20.5 (±8.7)	1 patient discontinued during 2nd day of hospital stay (excluded from all analyses)
Oncken USA 1996 [50]	Comparison of before–after cohorts	Pregnant smokers recruited through newspaper ads, TV announcements and from local obstetric and family practice clinics Inclusion and exclusion criteria specified. Data collection: screening physical examination, detailed smoking history taken. Then baseline maternal BP/HR, fetal HR and uterine/umbilical artery PVWs and resistance index (RI) measured 30 minutes after refraining from smoking, before and after smoking usual cigarette. Then randomly assigned to either NRT gum/	To evaluate short-term concentrations of nicotine delivered by nicotine gum use versus smoking and maternal and fetal haemodynamic parameters	36 women eligible for screening visit 7 excluded (5 for cotinine concentration too low, 2 urine tox. screen-positive.) 29 remaining Randomly assigned (2 : 1 assignment using simple randomization as anticipated not all gum users would complete study) to smoking versus chewing nicotine gum	NRT $n = 19$, stop smoking and chew gum containing 2 mg nicotine/piece At least 6 pieces per day and not more than 2 pieces/hour or 30 pieces/day

(*Continues*)

Table 2. (Continued)

Cohorts with NRT administered as an intervention in experimental settings						
First author, Location, Date	*Study design*	*Recruitment method (location/setting, eligibility criteria, data gathering)*	*Study objectives*	*Baseline characteristics/group differences*	*Intervention group(s) (including dose and route of NRT)*
		smoking and returned daily for monitoring, inc. maternal vital signs, adverse effects, withdrawal symptoms, CO measurements, smoking status and pieces of gum chewed. After 5 days, initial measurements repeated after 30 minutes abstaining			
Oncken USA 2009 [51]	Controlled cross-over cohort study	Recruitment: newspaper advertisements and flyers in local physician offices Inclusion and exclusion criteria specified. Data collection: screening visit for consent, medical history/physical exam/fetal ultrasound. Advised to reduce smoking to 10–15 cigarettes/day for ≥ 4 days. Session 1: after abstinence for ≥ 8 hours overnight, participants smoked an average of 1 cigarette of their choice/hour in negative pressure room with fetal monitoring. Baseline FHR measured before 1st and after 4th cigarette of day. Then randomly assigned to one of 3 groups (see intervention column). Session 2: on day 5, further monitoring similar to session 1 after 8 hours abstinence, with patch applied at 10 am and 11 doses nasal spray used	To examine the short-term effects of the nicotine patch or nasal spray on measures of nicotine exposure, withdrawal symptoms, and on maternal HR and FHR in pregnant smokers	29 subjects consented 8 did not complete whole study: 1 acted as a pilot subject, 3 did not attend first monitoring session, 2 had pregnancy complications before randomization, 2 dropped out after first day of medication Treatment for reasons unrelated to the study. Mean gestational age 31.26 (±2.61)	Women in study $n = 21$ $N = 7$ nicotine patch (15 mg/16 hours). Used for average 14 hours/day with placebo nasal spray $N = 7$ nasal spray (recommended regimen of 24 doses per day, each containing 1 mg nicotine) with placebo patch
Wright USA 1997 [52]	Uncontrolled before–after cohort study	Recruited from the University of North Carolina intervention in pregnancy study programme, all	To measure any short-term effects that transdermal NRT may have in pregnancy	6 women Mean maternal age 25.7 (21–31 years)	NRT $n = 6$, 21 mg nicotine patch applied for 6 hours

(*Continues*)

Table 2. (Continued)

Cohorts with NRT administered as an intervention in experimental settings

First author, Location, Date	*Study design*	*Recruitment method (location/setting, eligibility criteria, data gathering)*	*Study objectives*	*Baseline characteristics/group differences*	*Intervention group(s) (including dose and route of NRT)*
		recalcitrant to 'standard of care' for smoking cessation Inclusion and exclusion criteria specified. Admitted for total of 21 hours during which they agreed not to smoke cigarettes or chew gum. Patients continuously observed to ensure not smoking. Day 1, 9 p.m. (baseline): Maternal assessment (weight, BP, HR, RR, and temperature) and fetal assessment (biophysical profile and umbilical artery Doppler). Day 2, 8 a.m.: 21 mg nicotine patch administered. Fetal and maternal assessment repeated at 10 a.m. At 2 p.m., patch removed and further assessment performed		Mean gestational age 34.2 weeks (28.1–37.0 weeks) All reported smoking at least one half-pack per day but no more than two (CO verified)	
Outcomes: fetal effects and delivery outcomes after long-term NRT use in pregnancy					
Schroeder USA 2002 [53]	Within-patient, longitudinal before–after cohort study	Long term follow-up of patients in Ogburn Jr 1999 (see above for details of first part of study) After inpatient phase, subjects continued NRT patch for 8 weeks and returned weekly for follow-up until delivery. At each visit: non-stress testing of FHR, measurement of maternal vital signs and expired CO. Self-reported 7 day abstinence. Additional counselling provided for 5–10-minute sessions with review of personalized nicotine dependence	To describe smoking abstinence, fetal effects and delivery outcomes for pregnant smokers included in Ogburn Jr when treated with 8 weeks of patch therapy	21 pregnant women with singleton pregnancies	NRT $n = 21$, 22 mg/24-hour patch therapy Continued for 8 weeks Some patients discontinued patch therapy

(*Continues*)

Table 2. (Continued)

Cohorts with NRT administered as an intervention in experimental settings					
First author, Location, Date	*Study design*	*Recruitment method (location/setting, eligibility criteria, data gathering)*	*Study objectives*	*Baseline characteristics/group differences*	*Intervention group(s) (including dose and route of NRT)*
		treatment plan and self-help material. Estimated fetal weight also measured by ultrasound at 4 and 8 weeks. Followed up postnatally at 4 weeks, 6 months and 12 months			
Study reporting cohorts selected from intervention and usual care cohorts of quasi-randomized controlled trial with NRT as one component of study intervention					
Outcomes: pregnancy complications, preterm birth, small for gestational age					
Hegaard Denmark 2004 [54]	Cohort as part of larger quasi-randomized intervention study [55]	Pregnant smokers with uneven birth dates received individual counselling and were invited to join a smoking cessation course offering NRT. Those with even birth dates were offered usual care: routine information on risk of smoking in pregnancy and advice on cessation from midwives without specific cessation training. Women that accepted and suitable for NRT included as cohort intervention group. Comparator group formed by randomly selecting from those offered usual care. Study period November 1996–December 1999. Recruited from a large Copenhagen University Hospital at first prenatal visit. Inclusion/exclusion criteria specified	To describe the effectiveness and safety of NRT in 75 pregnant smokers, comparing with smoking	647 women included in the larger study in total (intervention $n = 327$, control $n = 320$) NRT group at baseline: 12.5 ± 5.2 (SD) cigarettes/day 21.5 ± 8.4 (SD) mean gestational age in weeks	NRT $n = 75$ $N = 25$ (33.3%) gum only $N = 31$ (41.3%) patch only $N = 19$ (25.3%) nicotine patch and gum NRT use was permitted for max. 11 weeks, participants not allowed to smoke. Dose and type based on Fagerström score. Smoking cessation training course provided alongside NRT by trained midwives

(*Continues*)

Table 2. (Continued)

Cohorts with NRT administered as an intervention in experimental settings

First author, Location, Date	*Comparator group(s)*	*Outcome measures (i.e. adverse fetal/maternal health outcomes)*	*Main/significant results*
Outcome: Fetal breathing movements during nicotine administration			
Gennser Sweden 1975 [43]	$N = 6$ tested with: ● Standard cigarette (1.7–1.8 mg nicotine) ● Non-tobacco cigarette (0.75–0.78 µg nicotine) Tested on 2 separate days	Fetal breathing movements—classified per 20 sec period as regular, irregular, periodic or apnoeic breathing patterns and frequency of each movement pattern given for each 5 minute period Maternal HR and RR	Maternal HR: HR significantly increased 5 minutes ($P < 0.001$) and 30 minutes ($P < 0.05$) after smoking standard cigarette, no significant change after smoking non-tobacco cigarette. After 2 mg nicotine gum, HR significantly increased at 30 minutes ($P < 0.05$), not significantly increased at 60 minutes. After 4 mg nicotine gum, HR significantly increased at 30 minutes and 60 minutes ($P < 0.01$). Maternal RR not influenced by any intervention. Fetal breathing movements: Significant increase in frequency of both periodic and apnoeic breathing (i.e. not regular breathing movements) after standard cigarette, max. change during 20–25 minute interval. Dose dependent but non-significant rises of apnoeic/periodic breathing movements was noted after both nicotine gums. No change after non-tobacco cigarette. Proportionally, irregular breathing movements decreased and regular breathing movements were unchanged. Changes independent of fetal age
Manning UK 1976 [44]	Other groups, not directly compared with NRT: smoking ×2 tobacco cigarettes $n = 47$ Smoking ×2 herbal cigarettes $n = 10$ (5 of these also smoked tobacco cigarettes)	Fetal chest wall movements *in utero* Proportion of time during which fetal breathing movements were present were measured at 5-minute intervals	6/7 women given 4 mg gum showed decrease in fetal breathing, 1 showed no difference, not a significant change overall 5/5 women given 8 mg nicotine gum showed decrease in fetal breathing. Mean proportion of time during which fetal breathing movements were present went from pre-chewing level of 75% (±4.93) to 48% (±10.6) at 25 minutes ($P < 0.05$) Gradual recovery noted—35 minutes after chewing started, proportion of fetal breathing movements not significantly different from pre-chewing
Outcome: fetal blood flow during nicotine administration			
Bruner USA 1991 [45]	Smoking, phase 1: smoking one cigarette containing 1.2 mg nicotine over 5 minutes	FVWs analysed by calculation of the ratio of the peak systolic excursion of the maximum velocity envelope to the diastolic trough (S : D ratio). Analysis on	$N = 35$ completed both phases $N = 12$ did not complete phase 2 (4 delivered baby, 5 declined further participation, 3 lost to follow-up)

(*Continues*)

Table 2. (Continued)

First author, Location, Date	*Comparator group(s)*	*Outcome measures (i.e. adverse fetal/maternal health outcomes)*	*Main significant results*
Cohorts with NRT administered as an intervention in experimental settings			
		average of 5 waveforms recorded from each vessel, insonation performed only during periods of fetal inactivity with apnoea	Significant increase in mean umbilical artery PVW S : D ratio (indicating possible increased fetal cardiac contractility) after smokers chewed gum ($P < 0.01$) and when those who smoked > 10/day smoked cigarette ($P < 0.05$) but no significant increase in uterine artery PVW S : D ratios. All changes returned to baseline by 10 minutes No significant difference between PVW S : D ratios for each group at baseline. Between PVW S : D ratios for non-smokers or prior smokers after phase 1 or 2 Between maternal or fetal HRs during any study period
Oncken USA 1997 [46]	Smoking $n = 15$ (same women), smoking *ad libitum* for 8 hours; mean cigarettes smoked: 9 ± 2	Maternal plasma nicotine concentrations; changes in the uterine and umbilical artery RIs; PVWs and RIs in the fetal middle cerebral artery; FHR, maternal BP and HR	Systolic and diastolic BP and maternal HR not significantly different between nicotine patch use/smoking. Significant time effects for systolic BP and maternal HR (both $P < 0.001$) with max. increases of 5 and 6 mmHg and 10–11 beats/minute 2 hours after baseline measurement in both groups. Diastolic BP also changed significantly over time in both groups ($P = 0.007$). The change in middle cerebral artery RI from baseline to 4 hours later was similar during patch use and smoking. There were no group differences in Doppler measurements of the middle cerebral, umbilical and uterine arteries. Time effects were significant for the middle cerebral artery RI ($P = 0.02$) and the uterine artery RI ($P = 0.02$). Baseline FHR reactivity changes were variable with no significant difference between groups and mean FHR was not significantly different between baseline and 4 hours for either group. No clinically significant adverse effects/pregnancy complications
Outcomes: maternal and fetal physiological observations during NRT administration			
Lehtovirta Finland 1983 [47]	Nicotine-free herbal cigarettes[a] $n = 16$, smoked herbal cigarette (Honeyrose de Luxe) for 5 minutes (8 in 2nd trimester, 8 in 3rd)	Baseline FHR; maternal BP and HR throughout; indices of FHR variability: interval index (II; standard deviation of fetal QRS intervals, long- term	In the 2nd trimester, the DI decreased significantly after 10 minutes chewing ($P < 0.001$), which lasted for 20 minutes after chewing. The II also decreased

(*Continues*)

Table 2. (Continued)

Cohorts with NRT administered as an intervention in experimental settings

First author, Location, Date	*Comparator group(s)*	*Outcome measures (i.e. adverse fetal/maternal health outcomes)*	*Main significant results*
	Different participants to NRT group	component of FHR variability) and the differential index (DI; SD of interval differences, short term component of FHR variability)	but only significantly for last 5 minutes of chewing ($P < 0.01$). Baseline FHR significantly increased at beginning of chewing ($P < 0.001$) and returned to pre-chewing level 5 minutes after chewing stopped. Maternal HR and diastolic BP increased a little during chewing, but systolic BP remained elevated for 20 minutes after chewing stopped (all significant, but *P*-variable for different time-points). In the 3rd trimester the DI decreased ($P < 0.001$), which lasted for 10 minutes after chewing. The DI did not change. Baseline FHR significantly decreased ($P < 0.001$), lasting for 10 minutes after chewing. Maternal HR, systolic and diastolic BP increased during chewing and returned to baseline 5–10 minutes after chewing (all significant, but *P*-variable for different time-points)
Lindblad Sweden 1987 [48]	Placebo chewing gum $n = 20$ (same women): chewed for 30 minutes	Maternal HR; maternal BP; FHR; fetal blood velocity; fetal vessel diameter	Nicotine gum significantly increased maternal HR from 5 to 45 minutes, systolic BP from 5 to 30 minutes and diastolic BP from 5 to 35 minutes, with variable *P*-values for different time-points. FHR and blood flow were not affected There was no change in the waveform of blood velocity in either the fetal aorta or the umbilical artery. One fetus had a supraventricular arrhythmia 15 minutes after gum chewing started which was sustained, but this had a normal heart rhythm the next day Mean gestational age at delivery 39.8 weeks (SD = 1.2). Mean birth weight 3424 g (SD = 445). All newborns had Apgar scores > 8 at 1 and 5 minutes. Mean umbilical arterial and venous pH 7.21 (SD = 0.08) and 7.30 (SD = 0.07) (within normal limits)

(*Continues*)

Table 2. (Continued)

Cohorts with NRT administered as an intervention in experimental settings			
First author, Location, Date	*Comparator group(s)*	*Outcome measures (i.e. adverse fetal/maternal health outcomes)*	*Main/significant results*
Ogburn Jr USA 1999 [49]	No comparator group for fetal/maternal observations except baseline measurements	Baseline measurements compared to inpatient phase for FHR, systolic: diastolic ratio in umbilical artery, maternal vital signs, maternal nicotine withdrawal scores	During days 2, 3 and 4 of inpatient phase, morning baseline FHR was significantly reduced relative to baseline when mother was smoking No significant changes in umbilical artery systolic/diastolic ratio from baseline at day 1 or day 4 No changes from baseline in maternal HR Reduction from baseline in overall maternal nicotine withdrawal score each morning
Oncken USA 1996 [50]	Smoking $n = 10$, continued smoking as usual	Maternal and fetal haemodynamic characteristics and nicotine/cotinine levels	No significant difference in any baseline characteristics. Mean maternal age in years ± SD: nicotine gum chewers 28 ± 6, cigarette smokers 27 ± 6. Mean gestational week at entry to study ± SD: nicotine gum chewers 28.1 ± 3.2, smokers 29.6 ± 3.6 Percentage change in haemodynamic parameters usually greater for smokers but when compared to gum, none significant 5 withdrew during study—4 from gum group: 2 unable to maintain abstinence on chewing gum, 2 severe nausea, 1 from smoking group due to acute bronchitis
Oncken USA 2009 [51]	$N = 7$ placebo nasal spray and patch	Nicotine concentration; maternal HR; FHR	Placebo and nicotine nasal spray groups saw a greater reduction in paired differences of the mean maternal HR when compared to session 1 than the nicotine patch group ($P = 0.021$). Baseline FHR decreased in the placebo group throughout session 2, but increased slightly in the patch and nasal spray groups. No serious adverse events during the study
Wright USA 1997 [52]	No comparator group	Maternal HR, BP, RR, ℓtemperature Fetal biophysical profile; FHR, amniotic fluid index Narrative birth complications	There were no measurable differences in fetal wellbeing during placement of patch. Maternal vital signs remained stable except for expected low maternal HR in morning after smoking cessation

(*Continues*)

Table 2. (Continued)

Cohorts with NRT administered as an intervention in experimental settings			
First author, Location, Date	*Comparator group(s)*	*Outcome measures (i.e. adverse fetal/maternal health outcomes)*	*Main/significant results*
			overnight, which gradually rose after patch applied. FHR baseline, decelerations and umbilical artery Doppler readings were unchanged. No fetus had significant changes in minute variation or accelerations and there were no changes in uterine activity. Ultrasonographic biophysical profiles remained unchanged Mean gestational age at delivery 40 weeks, mean birth weight 3239 g. All deliveries uncomplicated except for 1 which had deep variable decelerations and a low 1 minute Apgar score of 3; at 5 minutes Apgar was 9 5/6 patients denied any nicotine withdrawal symptoms. Symptoms reported: hunger, inadequate sleep, irritability, tobacco craving, impatience, restlessness, headache and drowsiness. All scores ≤ 5/10 on scale from 0 = absent to 10 = severe
Outcomes: fetal effects and delivery outcomes after long-term NRT use in pregnancy			
Schroeder USA 2002 [53]	No comparator group	Infant status at birth Any birth complications Details of remainder of the pregnancy	Only 8 patients finished study according to protocol: 1 withdrew on 2nd inpatient day. After inpatient phase: 7 withdrew within the 1st week, 3 after 5/52 and 2 after 6/52. 7 withdrew due to adverse events, 5 due to resuming smoking. Median time from start of patch therapy to delivery was 12 weeks. Centile weight for gestational age not found to change significantly during study. All weekly non-stress tests reactive initially or with extended observation. All 21 infants born alive. Mean gestational age at birth was 38.9 weeks, SD ± 1.3 (median 39.1), mean birth weight 3439 g, SD ± 570 (length, IIC. Apgars also reported). 3 suffered severe neonatal morbidity—1 fetal asystole and HIE, 1 complete transposition of great vessels, 1 mild RDS and seizures within 1 month

(*Continues*)

Table 2. (Continued)

First author, Location, Date	*Comparator group(s)*	*Outcome measures (i.e. adverse fetal/maternal health outcomes)*	*Main significant results*
Cohorts with NRT administered as an intervention in experimental settings			
Study reporting cohorts selected from intervention and usual care cohorts of quasi-randomized controlled trial with NRT as one component of study intervention			
Outcomes: pregnancy complications, preterm birth, small for gestational age			
Hegaard Denmark 2004 [54]	Smokers n = 150 received 'usual care' (see column 3). 2 smokers matched to each NRT user by daily cigarette consumption	Pregnancy related complications: abruption placentae; fetal death; preterm birth (< 37 weeks); small for gestational age	Of the NRT group; N = 30 (40%) used NRT for < 2 weeks; N = 28 (37%) for 2–6 weeks; N = 17 (23%) for 7–11 weeks Complications in pregnancy were similar in the NRT group to the smoking group. There were no fetal deaths in either group and one abruption placentae in the control group. Preterm birth: NRT n = 4 (5.3%); control n = 5 (3.3%) (P = 0.5). Small for gestational age: NRT n = 5 (6.7%); control n = 11 (7.3%) (P = 1.0)

All confidence intervals (CIs) 95% unless stated otherwise. 'Pregnant women prescribed or issued nicotine replacement therapy (NRT) were assumed to have smoked prior to that point in pregnancy; hence this group was exposed to both smoking and NRT. 'Pregnant women prescribed or issued NRT were assumed to have smoked prior to that point in pregnancy, so women who reported using NRT 'on its own' were pooled with other NRT users (who smoked concurrently) for analysis. Episodes of infantile colic in smokers were quoted as 11417 in a table in this paper but odds ratio (OR) suggests that this is a typographic error—adjusted to 1417. †Data on bupropion/herbal cigarette/varenicline use not displayed here. OTC = over-the-counter; PPROM = preterm premature rupture of membranes; ECG = electrocardiogram; CO = carbon monoxide; HR = heart rate; BP = blood pressure; LMP = labour management partnership; BMI = body mass index; FHR = fetal heart rate; HR = heart rate; BP = blood pressure; HIE = hypoxic–ischaemic encephalopathy; RR = respiratory rate; RRR = resting RR; FVWs = flow velocity waveforms; IQR = interquartile range; GP = general practitioner; MUMS Study = Maternal Use of Medications and Safety; OR = odds ratio.

Table 3 Summary of outcome measures by study: non-RCTs.

First author name and year of study	*Congenital anomalies*	*Low birth weight*	*Mean birth weight*	*Small for gestational age*	*Preterm birth*	*Mean gestational age at birth*	*Stillbirth*	*Fetal death*	*Mode of delivery*	*Birth outcomes*	*Infantile colic*	*Infant strabismus*	*Fetal observations during NRT administration*
Routine health-care studies													
Bérard 2016 [40]			✓	✓	✓	✓							
Dhalwani 2018 [32]							✓						
Dhalwani 2015 [34]	✓												
Dhalwani 2014 [38]		✓	✓		✓			✓	✓				
Gaither 2009 [36]		✓			✓								
Lassen 2010 [37]			✓		✓								
Mlhidou 2012 [39]		✓			✓						✓		
Morales-Suárez-Varela 2006 [35]	✓												
Strandberg-Larsen 2008 [33]							✓						
Torp-Pedersen 2010 [42]												✓	
Tran 2020 [41]				✓	✓				✓	✓			
Interventional studies													
Bruner 1991 [45]													✓
Genmser 1975 [43]													✓
Hegaard 2004 [54]				✓	✓					✓			
Lehtovirta 1983 [47]													✓
Lindblad 1987 [48]			✓			✓				✓			✓
Manning 1976 [44]													✓
Ogburn Jr 1999 [49]													✓
Oncken 1997 [46]													✓
Oncken 1996 [50]													✓
Oncken 2009 [51]													✓
Schroeder 2002 [53]			✓			✓				✓			✓
Wright 1997 [52]			✓			✓				✓			✓

NRT = nicotine replacement therapy.

Table 4 Modified Newcastle–Ottawa scale quality assessment scores: non-RCTs.

Routine health-care studies	*Selection (max. 5 stars)*	*Design and analysis (max. 1 star)*	*Outcome (max. 2 stars)*	*Total (max. 8 stars)*
Bérard 2016 [40]	★★★	★	★	★★★★★
Dhalwani 2018 [32]	★★★★	★	★	★★★★★★
Dhalwani 2015 [34]	★★★★	★	★	★★★★★★
Dhalwani 2014 [38]	★★★★	★	★	★★★★★★
Gaither 2009 [36]	★★	★	★	★★★★
Lassen 2010 [37]	★★★★	★	★★	★★★★★★★
Milidou 2012 [39]	★★★★	★		★★★★★
Morales-Suárez-Varela 2006 [35]	★★★★	★	★★	★★★★★★★
Strandberg-Larsen 2008 [33]	★★★★	★	★★	★★★★★★★
Torp-Pedersen 2010 [42]	★★★★	★	★★	★★★★★★★
Tran 2020 [41]	★★	★	❖	★★★★
Interventional studies	*Selection (max. 4 stars)*	*Design and analysis (max. 1 star)*	*Outcome (max. 2 stars)*	*Total (max. 7 stars)*
Bruner 1991 [45]	★	★	★★	★★★★
Gennser 1975 [43]	★	★	★★	★★★★
Hegaard 2004 [54]	★			★
Lehtovirta 1983 [47]			★	★
Lindblad 1987 [48]	★★	★	★	★★★★
Manning 1976 [44]	★		★	★★
Ogburn Jr 1999 [49]	★★	★	★★	★★★★★
Oncken 1997 [46]	★★	★	★★	★★★★★
Oncken 1996 [50]	★★		★★	★★★★
Oncken 2009 [51]	★★	★	★★	★★★★★
Schroeder 2002 [53]		★	★★	★★★
Wright 1997 [52]	★		★★	★★★

Quality assessment scores for routine health-care and interventional cohort studies as assessed by the modified Newcastle–Ottawa scale [20]; see Supporting information, Appendix S1 for scales. RCTs = randomized controlled trials.

[45,46,50], fetal breathing [43,44] and heart rate [43,46–48,50,51] and maternal blood pressure and heart rate [43,46–52]; some also reported pregnancy outcomes [48,52–54].

Quality assessment

Table 4 reports quality assessments. Routine health-care studies had a median score of 6/8 stars [interquartile range (IQR) = 5–7] and low scores often reflected a lack of validation of participants' exposures (e.g. NRT use), retrospective exposure assessment or a lack of adverse outcome validation. Interventional studies' median score was 4/7 stars (IQR = 2.5–4.5); these often scored poorly on cohort representativeness but relatively well for having biochemical validation of smoking abstinence.

Meta-analysis outcomes

We performed meta-analyses for congenital anomalies, stillbirth and preterm birth outcomes, but for others this was not possible due to differences in study designs. Analyses only included routine health-care studies. As interventional cohorts used 'before–after' designs without appropriate comparison groups, the few which reported birth outcomes could not be included. The study which investigated a subsample of quasi-RCT intervention group participants selected intervention and comparison groups in very different ways, and was judged unsuitable for inclusion [54].

Major congenital anomalies after first-trimester NRT exposure were reported using the European Surveillance of Congenital Anomalies and Twins (EUROCAT) classification system in two studies [34,35,56]. Stillbirth rate was reported in two; one study defined this as a baby born not showing signs of life at ≥ 28 weeks [32] and the other after 20 weeks [33]; we pooled these, as both represented death in later pregnancy. One interventional study reported fetal deaths but was excluded for the reason outlined above [54]. Preterm birth (at < 37 weeks) was an outcome in six studies, but only two were pooled [36,37]; three were without appropriate comparison groups [40,41,54] and one [39] duplicated findings from another included study [37].

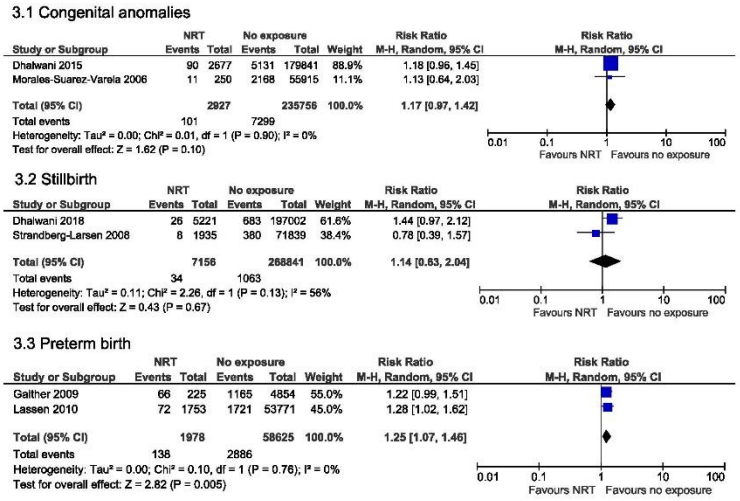

Figure 3 Meta-analyses of non-randomized controlled trials (RCTs). [Colour figure can be viewed at wileyonlinelibrary.com]

Meta-analysis results: non-RCTs

Figure 3 shows non-RCT meta-analysis findings. Compared with no NRT use, there was no evidence for an association between using NRT and risks of congenital anomalies (RR = 1.17, 95% CI = 0.97–1.42, I^2 = 0%; Fig. 3.1) or stillbirth (RR = 1.14, 95% CI = 0.63–2.04, I^2 = 56%; Fig. 3.2). Similarly, when compared to smoking, NRT use was not associated with anomalies (RR = 1.06, 95% CI = 0.86–1.32, I^2 = 0%) or stillbirth (RR = 0.75, 95% CI = 0.41–1.36, I^2 = 54%). Compared with no NRT use, meta-analysis of two studies suggested a slightly increased risk of preterm birth (RR = 1.25, 95% CI = 1.07–1.46, I^2 = 0%; Fig. 3.3) but, compared to smoking, NRT was not associated with greater preterm birth risk (RR = 1.12, 95% CI = 0.95–1.33, I^2 = 0%). For 'NRT versus no NRT' comparisons GRADE criteria certainty of evidence for these outcomes was 'very low'.

Narratively reported outcomes: non-RCTs

Table 3 reports outcomes by study. Two studies excluded from the preterm birth meta-analysis compared risks of preterm birth following NRT use in women who smoked; there was a significantly reduced risk in NRT users compared to non-users in one paper (adjusted OR = 0.21, 95% CI = 0.13–0.34) [40], while the second showed no significant difference (HR = 1.00, 95% CI = 0.71–1.42) [41].

Four studies reported mean gestational age at birth for NRT-exposed women [40,48,52,53] but only one, which enrolled only women who smoked, had a comparison group [40]; with no statistical comparison, this reported a mean (standard deviation (SD)) birth gestational age in NRT users of 38.9 (1.9) weeks and in non-NRT users of 37.5 (3.3).

Three studies reported small for gestational age (SGA) rates [40,41,54]. Two included only women who smoked, with one reporting a significantly reduced risk of SGA in those using NRT compared to those who did not (adjusted OR = 0.61, 95% CI = 0.41–0.90) [40] and the other showing no significant change in risk (HR = 0.77, 95% CI = 0.56–1.07) [41]. The other study used very different methods for selecting exposure groups rendering these non-comparable, but reported no significant difference in SGA rates [54].

Mean birth weight was reported by six studies [37,38,40,48,52,53]. three were interventional [48,52,53] and three had comparison groups which were too dissimilar to be aggregated [37,38,40]. One of these enrolled women who smoked reported, with no statistical comparison, a mean birth weight (SD) in NRT users of 3257.9 g (553.1) and non-users of 2943.5 g (733.5) [40]. A PhD thesis using medical record data compared mean birth weight in NRT users and women who neither smoked nor used NRT in pregnancy and found these were lower (β = −168 g, 99% CI = −214 to −122, $P < 0.001$)

Table 5 Summary of effects of NRT on fetal physiological outcomes: non-RCTs.

Study first author and year	*Participant smoking status at time of baseline measurements*	*Post-baseline exposures: NRT type and dose ± smoking*	*NRT effect on fetal observations*
Fetal breathing movements			
Gennser 1975 [43]	Abstinent for ≥12 hours	NRT gum 2 mg NRT gum 4 mg Smoking (1.7–1.8 mg nicotine cigarette)[c]	↔ Apnoeic/periodic breathing movements after both gums ↑ apnoeic/periodic breathing movements after tobacco cigarette
Manning 1976 [44]	Abstinent overnight	NRT gum 4 mg[d] NRT gum 8mg[d]	↔ Total breathing movements after 4 mg gum ↓ Fetal breathing movements after 8 mg gum
Fetal blood flow			
Bruner 1991[a] [45]	Abstinent for ≥ 30 minutes	NRT gum 2 mg smoking (1.2 mg nicotine cigarette)[c]	↑ Umbilical artery flow velocity waveform (FVW) systolic: diastolic (S : D) ratios when all current smokers chewed gum ↑ Umbilical artery FVW S : D ratios when those who smoked > 10/day smoked a cigarette ↔ Uterine artery FVW S : D ratios after current smokers chewed gum ↔ Uterine artery FVW S : D ratios after current smokers smoked a cigarette ↔ Umbilical artery or uterine artery FVW S : D ratios when prior smokers/non-smokers chewed gum ↔ Umbilical artery or uterine artery FVW S : D ratios when prior smokers/non-smokers smoked a cigarette
Oncken 1997 [46]	Abstinent from 8 p.m. the day before the study	NRT patch 21 mg/applied for 8 hours. Smoking cigarettes *ad libitum* for 8 hours[c]	↔ Change from baseline in middle cerebral artery resistance index after patch compared to change after smoking ↔ Change from baseline FHR reactivity after patch compared to after smoking
Fetal physiological observations			
Lehtovirta 1983[b] [47]	Not stated in Methods	NRT gum 2 mg (chewed by women in 2nd and 3rd trimester)	↓ Differential index[e] after 10 minutes of chewing gum in 2nd trimester, ↔ in 3rd trimester ↓ interval index[f] in both 2nd (last 5 minutes of chewing) and 3rd trimesters ↑ FHR at start of chewing in 2nd trimester ↓ FHR in 3rd trimester
Lindblad 1987 [48] Ogburn Jr 1999 [49]	Abstinent for ≥ 2 hours Abstinent in smoke-free setting for 4 days	NRT gum 4 mg NRT 22 mg/24 hour patch	↔ FHR or in waveform of blood velocity in fetal aorta or umbilical artery ↓ Baseline FHR in mornings
Oncken 1996 [50]	Abstinent for ≥ 30 minutes	NRT 2 mg gum[d]. Smoking[d]	↔ In umbilical artery systolic/diastolic ratio ↔ Percentage change from baseline in FHR or umbilical/uterine artery flow velocity waveforms after NRT compared to after smoking. Between-group differences reported

(*Continues*)

Table 5. (Continued)

Study, first author and year	*Participant smoking status at time of baseline measurements*	*Post-baseline exposures: NRT type and dose ± smoking*	*NRT effect on fetal observations*
Oncken 2009 [51]	Abstinent for ≥ 8 hours	NRT patch 15 mg/16 hours[d] Nasal spray 24 doses per day 1 mg/dose[d]	↔ FHR in either patch or nasal spray groups compared to placebo
Schroeder 2002 [53]	Smoking alongside NRT use	NRT patch 22 mg/24 hour (over 8 weeks)	Weekly non-stress tests all reactive
Wright 1997 [52]	Abstinent overnight in supervised setting	NRT patch 21 mg applied for 6 hours while abstinent	↔ Baseline FHR, decelerations, umbilical artery Doppler readings or biophysical profile

↔ No significant change; ↑ Significant increase; ↓ Significant decrease.
All participants in these studies were smokers at the start of the study, unless otherwise stated in footnotes.
[a]Participants were: non-smokers, prior smokers, smokers < 10 cigarettes/day, smokers > 10 cigarettes/day.
[b]Smoking status of nicotine replacement therapy (NRT)-exposed women unclear from Methods.
[c]Exposures sequentially in one group of women.
[d]Exposures in different groups of women.
[e]Differential index: short-term component of fetal heart rate (FHR) variability; lower values suggest reduced fetal blood flow [47].
[f]Interval index: long-term component of FHR variability, lower values suggest reduced fetal blood flow.

[38]. Within a multivariate analysis which adjusted for reported smoking behaviour, a population-based cohort found no statistically significant associations between duration of NRT use and mean birth weight (β = 0.25 g per week of NRT use, CI = −2.31 to 2.81) [37].

Low birth weight (less than 2500 g) was reported by three studies which seemed similar enough to be aggregated, but due to heterogeneity (I^2 = 76%) are presented separately [36,38,39]. One reported low birth weight incidences of 2.4% in unexposed women, 2.9% in NRT users, 4.8% of women who smoked and used NRT and 4.3% in smokers [39]. A retrospective questionnaire study found that 13.1% of NRT-exposed women delivered low birth weight infants and rates were 9.26% within women who smoked and 6.99% with neither exposure [36]. Another study reported that NRT exposure was associated with increased risk of low birth weight when compared to no exposure (OR = 1.88, 99% CI = 1.42–2.49, $P < 0.001$) [38]. Two of these studies had the lowest quality scores of all routine health-care studies (see Table 4) [36,39].

Fetal death, a composite of stillbirth and miscarriage [38], delivery mode [38], infantile colic [39] and infant strabismus [42], were reported in single studies and Table 2 reports these findings. Compared with no NRT use, exposure was associated with reduced risk of fetal death (OR = 0.44, 99% CI = 0.38–0.50, $P < 0.001$) [38] and of assisted delivery (relative RR (RRR) = 0.68, 99% CI = 0.54–0.85, $P < 0.001$) but not with increased risk of caesarean section [38]. A study of women who smoked who were exposed to NRT reported a composite outcome: 'any adverse perinatal event', encompassing a number of separate birth outcomes [41]. Table 2 reports the individual outcome HRs, but there was no significant change in overall risk of any adverse perinatal event when comparing women who smoked who were exposed to NRT and those who were not (HR = 1.02, 95% CI = 0.84–1.23).

Table 5 presents physiological outcomes measured by study. In nine studies, fetal physiological observations were recorded at baseline and compared to readings taken when abstinent and using NRT [43–46,48–52]. Three also compared these within-patient changes from baseline with those recorded during or after smoking following a similar period of abstinence [43,45,46]. Results showed no consistent patterns, and most studies did not report significant outcome changes after NRT administration.

DISCUSSION

Key findings

Overall, we found no evidence that NRT used by pregnant women who smoke has adverse impacts on fetal and infant outcomes. Although underpowered, the direction of point-estimates derived from most RCT meta-analyses

suggest that NRT is not likely to have adverse impacts or be more harmful than smoking in pregnancy. The robustness of non-RCT evidence was poor, with meta-analyses' findings affected by imprecision or potential biases, which may explain the inconsistency in the direction of associations found in non-RCT meta-analyses. NRT-exposed women are likely to have smoked at some point in pregnancy but, generally, this was not measured and so could not be adjusted for in non-RCTs, making interpretation of these studies' findings particularly difficult.

Strengths and limitations

Our synthesis meta-analyses of non-RCT studies are limited by the inherent biases in these study designs. An issue was that ascertainment of NRT exposure relied upon maternal self-report or prescription records. Women's recall may not have been perfect and, as some women prescribed NRT will not have used it, using prescription records could overestimate NRT exposure. More importantly, studies generally assessed NRT exposure at only one or two time-points in pregnancy and in most, smoking intensity either before or after NRT use was not reported, despite smoking being known to adversely affect outcomes. The omission of detailed smoking data from non-RCT reports was probably the greatest threat to these studies' validity. It is logical to assume that all women issued NRT would have smoked at least in early pregnancy, and this will have tended to reduce differences between exposure groups' outcomes. Only two non-RCT studies adjusted for smoking behaviour [33,37]; others could be subject to confounding of unknown magnitude. Another important issue was that NRT prescribing involved confounding by indication [57]. In three of the five birth cohorts which provided non-RCT studies' data, women issued with NRT had higher rates of comorbidities and lower socio-economic status than other women who smoked, and so very probably experienced 'higher-risk' pregnancies [10,33,36] which may have substantially affected adverse outcomes. We believe that our modified NOS for non-RCTs' quality assessments and the application of GRADE criteria should help readers to understand the degree to which observed associations might be causal or due to bias, confounding or chance.

For the non-RCT review, only one reviewer screened titles and abstracts and extracted data; although another person checked this, there was no parallel independent screening or extraction by the second researcher, so researcher bias is a possibility. Additionally, some non-RCTs may not have been indexed in databases, but we are confident that our comprehensive search strategy will have found all which were and, hopefully, methods for assessing bias and certainty of non-RCT evidence assist the findings' interpretation.

Strengths of this work include applying 'Cochrane-type' review methods to find all available and relevant RCTs and non-RCTs. We believe this is the first attempt to systematically retrieve and synthesize all studies which report fetal and infant health outcomes after pregnant women have used or been offered NRT, and that we have successfully identified, assessed and presented together all relevant studies. This, coupled with objective methods for assessing studies' biases and the strength of evidence produced by meta-analyses, should provide a thorough report of what is known about the impact of NRT on pregnancy outcomes. Similar reviews have had less thorough search strategies, presented only narrative data or have not attempted to assess bias [15,58,59]. While meta-analyses are underpowered, these remain the strongest currently available data on NRT safety in pregnancy, and strengths and weaknesses of the literature are highlighted. The juxtaposition of non-RCT and RCT meta-analyses is perhaps the most useful feature of the review, and is illustrated by considering findings regarding preterm birth. For this outcome, meta-analysis of two non-RCT studies revealed a statistically significant association between NRT use and higher rates of prematurity in which we have 'very low' certainty. However, meta-analysis of data from seven RCTs provides a non-statistically significant 'best estimate' for this association being in the opposite (protective) direction. This direct comparison helps the reader to more clearly appreciate and consider the quality of available data before drawing conclusions. This disparity might be explained by women's smoking either before, after or alongside NRT exposure, which was generally not adjusted for by non-RCTs. Smoking is well known to contribute to increased risk of pre-term birth [60], and one of the included studies in this meta-analysis acknowledges that the women recommended or prescribed NRT by a health-care professional might be those who smoke more heavily [36] and find it harder to quit [61].

Findings in context of previous literature

The most robust research on the safety of NRT in pregnancy comes from RCTs, and we report meta-analyses for nine safety-orientated outcomes [13]. In RCTs there is no confounding by indication, and randomization ensures that unknown confounders are distributed equally between trial groups, so differences in birth outcomes can be assumed to be caused by NRT. Although meta-analyses were underpowered and there were no significant differences between the NRT and control groups, the trend in non-statistically significant point estimates derived from these analyses is noteworthy. For low birth weight, preterm birth, neonatal intensive care unit admissions, neonatal death and congenital anomalies, point estimates suggest a protective effect of NRT, whereas those for miscarriage

and stillbirth do not. Additionally, caesarean section rates were non-significantly higher following NRT but, in the absence of contextual data, it is not clear if this is an adverse or a positive outcome. This point estimate trend suggests that, with more data from RCTs, NRT could well prove to be less harmful than smoking in pregnancy. Due to design issues, non-RCT meta-analyses are probably not methodologically robust enough to inform clinical practice and their findings do not add to those from RCT meta-analyses. Pregnant women in non-RCT studies are only likely to have been prescribed or offered NRT by clinicians if they smoked. Consequently, to provide valid findings, these studies should have assessed pregnant women's smoking behaviour and adjusted analyses for this. As the probable mechanism for NRT improving birth outcomes is due to women stopping smoking or smoking less, this is particularly important.

Further work

RCTs and robust population-based cohort studies from routine health-care settings are needed to improve the evidence base for the safety of NRT use in pregnancy. Electronic medical records databases offer the potential for valid capture of near-complete pregnancy outcome data. However, to make a valid contribution to the literature, future non-RCT studies need better methods for quantifying exposures to NRT and smoking during the whole of pregnancy and to adjust for the latter in analyses.

CONCLUSIONS

The strongest data on the probable impacts of NRT exposure in pregnancy on birth outcomes comes from RCTs, and these provide no suggestion that NRT might be harmful. Non-RCT studies have less consistent findings, due most probably to inherent design weaknesses, and future observational studies should provide analyses which account for the impact of smoking behaviour within women who also use NRT in pregnancy.

Declaration of interests

None.

Acknowledgements

This project was funded through a National Institute for Health Research (NIHR) Programme Grants for Applied Research (Programme number RP-PG-0615-20003). T. C. is a National Institute for Health Research (NIHR) Senior Investigator. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Author contributions

Lauren Taylor: Data curation; formal analysis; investigation; methodology; project administration; validation; visualization. **Ravinder Claire:** Formal analysis; validation; visualization. **Katarzyna Campbell:** Formal analysis; methodology; validation. **Tom Coleman-Haynes:** Data curation; validation. **Sue Cooper:** Conceptualization; funding acquisition; methodology; supervision. **Tim Coleman:** Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; supervision; validation.

References

1. Batstra L., Hadders-Algra M., Neeleman J. Effect of antenatal exposure to maternal smoking on behavioural problems and academic achievement in childhood: prospective evidence from a Dutch birth cohort. *Early Hum Dev* 2003; **75**: 21–33.
2. DiFranza JR, Lew RA. Effect of maternal cigarette smoking on pregnancy complications and sudden infant death syndrome. J Fam Pract 1995; 40: 385–394.
3. Hackshaw A., Rodeck C., Boniface S. Maternal smoking in pregnancy and birth defects: a systematic review based on 173 687 malformed cases and 11.7 million controls. *Hum Reprod Update* 2011; **17**: 589–604.
4. Statistics Team NHS Digital. Statistics on Women's Smoking Status at Time of Delivery: NHS digital. Government Statistical Service; 2019. Available at: https://docs.google.com/viewer?url=https%3A%2F%2Ffiles.digital.nhs.uk%2FE7%2F4C5AAB%2Fstat-wome-smok-time-deli-eng-q3-18-19-rep.pdf (accessed 26 March 2019).
5. National Health Service Information Centre. *Infant Feeding Survey 2010*. Early results. London, UK: National Health Service Information Centre for Health and Social Care; 2011.
6. Gray R., Bonellie S. R., Chalmers J., Greer I., Jarvis S., Kurinczuk J. J., *et al.* Contribution of smoking during pregnancy to inequalities in stillbirth and infant death in Scotland 1994-2003: retrospective population based study using hospital maternity records. *BMJ* 2009; **339**: b3754.
7. Chamberlain C., O'Mara-Eves A., Oliver S., Caird J. R., Perlen S. M., Eades S. J., *et al.* Psychosocial interventions for supporting women to stop smoking in pregnancy. *Cochrane Database Syst Rev* 2013; **10**: CD001055.
8. Vaz L. R., Jones M. J., Szatkowski L., Tata L. J., Petrou S., Coleman T. Estimating the health-care costs of children born to pregnant smokers in England: cohort study using primary and secondary health-care data. *Addiction* 2018; 113: 1305–16.
9. National Institute for Health and Clinical Excellence (NICE). Public Health Guidance 26: How to Stop Smoking in Pregnancy and Following Childbirth. Report no.: PH26. London: NICE; 2010.
10. Dhalwani N. N., Szatkowski L., Coleman T., Fiaschi L., Tata L. J. Prescribing of nicotine replacement therapy in and around pregnancy: a population-based study using primary care data. *Br J Gen Pract* 2014; **64**: e554–e560.
11. Fish L. J., Peterson B. L., Namenek Brouwer R. J., Lyna P., Oncken C. A., Swamy G. K., *et al.* Adherence to nicotine replacement therapy among pregnant smokers. *Nicotine Tob Res* 2009; **11**: 514–8.

12. Hollands G. J., Sutton S., McDermott M. S., Marteau T. M., Aveyard P. Adherence to and consumption of nicotine replacement therapy and the relationship with abstinence within a smoking cessation trial in primary care. *Nicotine Tob Res* 2013; 15: 1537–44.
13. Claire R., Chamberlain C., Davey M. A., Cooper S. E., Berlin I., Leonardi-Bee J. *et al.* Pharmacological interventions for promoting smoking cessation during pregnancy. *Cochrane Database Syst Rev* 2020; 3: CD010078.
14. Bowker K., Campbell K. A., Coleman T., Lewis S., Naughton F., Cooper S. Understanding pregnant smokers' adherence to nicotine replacement therapy during a quit attempt: a qualitative study. *Nicotine Tob Res* 2016; **18**: 906–12.
15. Dempsey D. A., Benowitz N. L. Risks and benefits of nicotine to aid smoking cessation in pregnancy. *Drug Saf* 2001; **24**: 277–322.
16. Taylor L., Coleman, T., Coleman-Haynes, T., Campbell, K., Cooper, S., Dickinson, A. Systematic review of observational studies to compare the safety of nicotine replacement therapy (NRT) use and smoking in pregnancy. PROSPERO 2018 CRD42018085834.
17. Moher D., Liberati A., Tetzlaff J., Altman D. G., The P. G. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLOS Med* 2009; 6: e1000097.
18. Coleman T., Chamberlain C., Davey M. A., Cooper S. E., Leonardi-Bee J. Pharmacological interventions for promoting smoking cessation during pregnancy. *Cochrane Database Syst Rev* 2015; **12**: CD010078.
19. Covidence systematic review software: Veritas health innovation; 2018. Available at: www.covidence.org (accessed 27 March 2019).
20. Wells G., Shea B., O'Connell D., Peterson J., Welch V., Losos M. *et al.* The Newcastle–Ottawa scale (NOS) for assessing the quality of non-randomized studies in meta-analysis. 2000. Available at: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp (accessed 27 March 2019).
21. Mantel N., Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959; **22**: 719–48.
22. Schünemann H. B. J., Guyatt G., Oxman A. GRADE handbook for grading quality of evidence and strength of recommendations. 2013 [updated October 2013. Available at:; https://gradepro.org/cite/guidelinedevelopment.org/handbook (accessed 9 June 2019).
23. Berlin I., Grangé G., Jacob N., Tanguy M. L. Nicotine patches in pregnant smokers: randomised, placebo controlled, multicentre trial of efficacy. *BMJ* 2014; **348**: g1622.
24. Coleman T., Cooper S., Thornton J. G., Grainge M. J., Watts K., Britton J., *et al.* A randomized trial of nicotine-replacement therapy patches in pregnancy. *N Engl J Med* 2012; **366**: 808–18.
25. El-Mohandes A. A. E., Windsor R., Tan S., Perry D. C., Gantz M. G., Kiely M. A randomized clinical trial of trans-dermal nicotine replacement in pregnant African-American smokers. *Matern Child Health J* 2013; **17**: 897–906.
26. Oncken C., Dornelas E., Greene J., Sankey H., Glasmann A., Feinn R. *et al.* Nicotine gum for pregnant smokers: a randomized controlled trial. *Obstet Gynecol* 2008; 112: 859–67.
27. Oncken C., Dornelas E. A., Kuo C.-L., Sankey H. Z., Kranzler H. R., Mead E. L., *et al.* Randomized trial of nicotine inhaler for pregnant smokers. *Am J Obstet Gynecol* 2019; **1**: 10–8.
28. Pollak K. I., Oncken C. A., Lipkus I. M., Lyna P., Swamy G. K., Pletsch P. K., *et al.* Nicotine replacement and behavioral therapy for smoking cessation in pregnancy. *Am J Prev Med* 2007; **33**: 297–305.
29. Wisborg K., Henriksen T. B., Jespersen L. B., Secher N. J. Nicotine patches for pregnant smokers: a randomized controlled study. *Obstet Gynecol* 2000; **96**: 967–71.
30. Cooper S., Taggar J., Lewis S., Marlow N., Dickinson A., Whitemore R. *et al.* Effect of nicotine patches in pregnancy on infant and maternal outcomes at 2 years: follow-up from the randomised, double-blind, placebo-controlled SNAP trial. *Lancet Respir Med* 2014; 2: 728–37.
31. TE Squires J., Bricker D., Potter L. *ASQ-3 User's Guide for the Ages and Stages Questionnaires, 3rd edn.* Baltimore: Paul H. Brookes Publishing Co; 2009.
32. Dhalwani NN, Szatkowski L, Coleman T, Fiaschi L, Tata LJ. Stillbirth among women prescribed nicotine replacement therapy in pregnancy: analysis of a large UK pregnancy cohort. *Nicotine Tob Res.* 2019; 21: 409–415.
33. Strandberg-Larsen K., Tinggaard M., Nybo Andersen A. M., Olsen J., Grønbæk M. Use of nicotine replacement therapy during pregnancy and stillbirth: a cohort study. *Br J Obstet Gynaecol* 2008; **115**: 1405–10.
34. Dhalwani N. N., Szatkowski L., Coleman T., Fiaschi L., Tata L. J. Nicotine replacement therapy in pregnancy and major congenital anomalies in offspring. *Pediatrics* 2015; **135**: 859–67.
35. Morales-Suarez-Varela M. M., Bille C., Christensen K., Olsen J. Smoking habits, nicotine use, and congenital malformations. *Obstet Gynecol* 2006; **107**: 51–7.
36. Gaither K. H., Brunner Huber L. R., Thompson M. E., Huet-Hudson Y. M. Does the use of nicotine replacement therapy during pregnancy affect pregnancy outcomes? *Matern Child Health J* 2009; **13**: 497–504.
37. Lassen T. H., Madsen M., Skovgaard L. T., Strandberg-Larsen K., Olsen J., Andersen A.-M. N. Maternal use of nicotine replacement therapy during pregnancy and offspring birthweight: a study within the Danish National Birth Cohort. *Paediatr Perinat Epidemiol* 2010; **24**: 272–81.
38. Dhalwani N. N. *Using primary care data to assess population-level estimates of maternal smoking and nicotine replacement therapy during pregnancy*; 2014. PhD thesis. University of Nottingham, UK.
39. Milidou I., Henriksen T. B., Jensen M. S., Olsen J., Søndergaard C. Nicotine replacement therapy during pregnancy and infantile colic in the offspring. *Pediatrics* 2012; **129**: e652–e658.
40. Bérard A., Zhao J.-P., Sheehy O. Success of smoking cessation interventions during pregnancy. *Am J Obstet Gynecol* 2016; **215**: 611.e1–8.
41. Tran D. T., Preen D. B., Einarsdottir K., Kemp-Casey A., Randall D., Jorm L. R., *et al.* Use of smoking cessation pharmacotherapies during pregnancy is not associated with increased risk of adverse pregnancy outcomes: a population-based cohort study. *BMC Med* 2020; **18**: 15.
42. Torp-Pedersen T., Boyd H. A., Poulsen G., Haargaard B., Wohlfahrt J., Holmes J. M., *et al. In-utero* exposure to smoking, alcohol, coffee, and tea and risk of strabismus. *Am J Epidemiol* 2010; **171**: 868–75.
43. Gennser G., Maršál K., Brantmark B. Maternal smoking and fetal breathing movements. *Am J Obstet Gynecol* 1975; **123**: 861–7.
44. Manning F. A., Feyerabend C. Cigarette smoking and fetal breathing movements. *Br J Obstet Gynaecol* 1976; **83**: 262–70.

45. Bruner J. P., Forouzan I. Smoking and buccally administered nicotine. Acute effect on uterine and umbilical artery Doppler flow velocity waveforms. *J Reprod Med* 1991; **36**: 435–40.
46. Oncken C. A., Hardardottir H., Hatsukami D. K., Lupo V. R., Rodis J. F., Smeltzer J. S. Effects of transdermal nicotine or smoking on nicotine concentrations and maternal–fetal hemodynamics. *Obstet Gynecol* 1997; **90**: 569–74.
47. Lehtovirta P., Forss M., Rauramo L., Kariniemi V. Acute effects of nicotine on fetal heart rate variability. *BJOG* 1983; **90**: 710–5.
48. Lindblad A., Marsal K. Influence of nicotine chewing gum on fetal blood flow. *J Perinat Med* 1987; **15**: 13–9.
49. Ogburn P. L. Jr., Hurt R. D., Croghan I. T., Schroeder D. R., Ramin K. D., Offord K. P., *et al.* Nicotine patch use in pregnant smokers: nicotine and cotinine levels and fetal effects. *Am J Obstet Gynecol* 1999; **181**: 736–43.
50. Oncken C. A., Hatsukami D. K., Lupo V. R., Lando H. A., Gibeau L. M., Hansen R. J. Effects of short-term use of nicotine gum in pregnant smokers. *Clin Pharmacol Ther* 1996; **59**: 654–61.
51. Oncken C., Campbell W., Chan G., Hatsukami D., Kranzler H. R. Effects of nicotine patch or nasal spray on nicotine and cotinine concentrations in pregnant smokers. *J Matern Fetal Neonatal Med* 2009; **22**: 751–8.
52. Wright L. N., Thorp J. M. Jr., Kuller J. A., Shrewsbury R. P., Ananth C., Hartmann K. Transdermal nicotine replacement in pregnancy: maternal pharmacokinetics and fetal effects. *Am J Obstet Gynecol* 1997; **176**: 1090–4.
53. Schroeder D. R., Ogburn P. L. Jr., Hurt R. D., Croghan I. T., Ramin K. D., Offord K. P., *et al.* Nicotine patch use in pregnant smokers: smoking abstinence and delivery outcomes. *J Matern Fetal Neonatal Med* 2002; **11**: 100–7.
54. Hegaard H. K., Kjærgaard H., Møller L. F., Wachmann H., Ottesen B. Long-term nicotine replacement therapy. *Br J Midwifery* 2004; **12**: 214–20.
55. Hegaard H. K., Kjaergaard H., Moller L. F., Wachmann H., Ottesen B. Multimodal intervention raises smoking cessation rate during pregnancy. *Acta Obstet Gynecol Scand* 2003; **82**: 813–9.
56. European Registration of Congenital Anomalies and Twins (EUROCAT). 'Guide 1.3' 2011. Available at: http://www.eurocat-network.eu/ABOUTUS/DataCollection/GuidelinesforRegistration/Guide1_3InstructionManual (accessed 11 December 2018).
57. Psaty B. M., Koepsell T. D., Lin D., Weiss N. S., Siscovick D. S., Rosendaal F. R., *et al.* Assessment and control for confounding by indication in observational studies. *J Am Geriatr Soc* 1999; **47**: 749–54.
58. Kreyberg I., Nordhagen I. S., Bains K. E. S., Alexander J., Becher R., Carlsen K. H., *et al.* An update on prevalence and risk of snus and nicotine replacement therapy during pregnancy and breastfeeding. *Acta Paediatr* 2019; **108**: 1215–21.
59. Bar-Zeev Y., Lim L. L., Bonevski B., Gruppetta M., Gould G. S. Nicotine replacement therapy for smoking cessation during pregnancy. *Med J Aust* 2018; **208**: 46–51.
60. National Center for Chronic Disease Prevention, Health Promotion Office on Smoking and Health *Reports of the Surgeon General. The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General.* Atlanta, GA: Centers for Disease Control and Prevention; 2014.
61. Riaz M., Lewis S., Naughton F., Ussher M. Predictors of smoking cessation during pregnancy: a systematic review and meta-analysis. *Addiction* 2018; **113**: 610–22.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Example search – used in Medline database
Appendix S1. Quality assessment scales and modification

Using Trial Sequential Analysis for estimating the sample sizes of further trials: example using smoking cessation intervention – Pre-print

Using Trial Sequential Analysis for estimating the sample sizes of further trials: example using smoking cessation intervention

Ravinder Claire (ravinder.claire@nottingham.ac.uk)
University of Nottingham https://orcid.org/0000-0001-8039-6161
Christian Gluud
Rigshospitalet
Ivan Berlin
Hopital Pitie-Salpetriere Service de Pharmacologie
Tim Coleman
University of Nottingham School of Medicine
Jo Leonardi-Bee
University of Nottingham Division of Epidemiology and Public Health

Research article

Keywords: Meta-analysis, trial sequential analysis methods, Trial Sequential Analysis software, sample size, information size, smoking, pregnancy, randomised clinical trial, pilot trial, feasibility trial

DOI: https://doi.org/10.21203/rs.3.rs-35669/v2

Abstract

Background: Assessing benefits and harms of health interventions is resource-intensive and often requires feasibility and pilot trials followed by adequately powered randomised clinical trials. Data from feasibility and pilot trials are used to inform the design and sample size of the adequately powered randomised clinical trials. When a randomised clinical trial is conducted, results from feasibility and pilot trials may be disregarded in terms of benefits and harms.

Methods: We describe using feasibility and pilot trial data in the Trial Sequential Analysis software to estimate the required sample size for one or more trials investigating a behavioural smoking cessation intervention. We show how data from a new, planned trial can be combined with data from the earlier trials using trial sequential analysis methods to assess the intervention's effects.

Results: We provide a worked example to illustrate how we successfully used the Trial Sequential Analysis software to arrive at a sensible sample size for a new randomised clinical trial and use it in the argumentation for research funds for the trial.

Conclusions: Trial Sequential Analysis can utilise data from feasibility and pilot trials as well as other trials, to estimate a sample size for one or more, similarly designed, future randomised clinical trials. As this method uses available data, estimated sample sizes may be smaller than they would have been using conventional sample size estimation methods.

Background

Demonstrating that health interventions work requires substantial resources. Often feasibility and pilot randomised clinical trials (RCTs) are conducted before larger-scale randomised clinical trials are designed to determine benefits and harms (1-3). Feasibility trials are used to ascertain information such as intervention acceptability, feasibility of intervention delivery, and recruitment likelihood to help design more decisive RCTs (1). A pilot trial is a smaller version of a large-scale RCT, and is used to test whether the main components of the trial, such as recruitment, randomisation, treatment, and follow-up assessments can all work together (1). Moreover, their data can be used to inform sample sizes for large-scale RCTs (2, 3).

Trial sequential analysis is a methodology that can be used in systematic reviews and meta-analyses to control random errors, and to assess whether further trials need to be conducted (4, 5). Trial sequential analysis as a method can be performed using the Trial Sequential Analysis software, which is freely available alongside its user manual online at The Copenhagen Trial Unit website (6). Here we employ Trial Sequential Analysis and combine data from feasibility and pilot RCTs testing a text message-based smoking cessation intervention for pregnant women ('MiQuit') (7, 8) to estimate the sample size that one or more future RCTs would need to recruit, to provide a more decisive answer regarding the effect of the intervention. We also show how data from the new, planned trial or trials can be combined with data from earlier trials using Trial Sequential Analysis to assess the intervention's benefits and harms. Using Trial Sequential Analysis sample size estimation methods maximises use of available trial data and consequently, the new RCT or trials may become smaller than they would have been using conventional sample size estimation methods.

Conventional meta-analysis

Meta-analyses often influence future research; when planning future trials, investigators frequently use meta-analysis to provide an accurate summary of an intervention's likely effect. If all available RCTs are included, systematic reviews with meta-analyses are considered the best available evidence, because power and precision of the estimated intervention effect is the best one can get (9, 10). However, this does not necessarily mean that the available evidence is either sufficient or strong. Conventional meta-analysis methods do not consider the amount of the available evidence in relation to the required sample size (11-13). The reliability of a statistically significant intervention effect generated by meta-analysis is often overvalued, particularly where sparse data (number of events and participants) or repetitive analyses (type I errors) are employed (6, 10, 14, 15). In other situations, intervention effects that are not statistically significant are often interpreted as showing that the intervention has no effect, and it is assumed that no more evidence is required (type II errors) (16, 17).

In conventional meta-analysis, there is no way to differentiate between an underpowered meta-analysis and a true finding of an intervention being 'ineffective'. However, it is imperative that a conclusion as to whether an intervention is truly ineffective or truly effective is made as soon as possible after trials are completed, in order to guide investigators' decisions as to whether further trials could be informative or not (6). Trial sequential analysis is a methodology that can overcome this issue by distinguishing whether meta-analyses provide evidence for either beneficial or harmful intervention effects, lack of effect (futility), or insufficient evidence for evaluation of the intervention effect (6, 18).

Methods

Trial Sequential Analysis

Meta-analyses aim to discover the benefit or harm of an intervention as early and as reliably as possible. As a result, they tend to be updated when new trials are published (19). When intervention evaluation has just begun and only few, smaller trials are available, meta-analyses may be conducted on sparse amounts of data and are at high risk of random type I and type II errors (20). As meta-analyses are updated they are subjected to repeated significance testing, which increases the risk of type I errors (21). When there are few data available, the Trial Sequential Analysis software resolves these issues by having stringent thresholds for assessing statistical significance, using monitoring boundaries. Monitoring boundaries also take into account the volume of significance testing which has been undertaken through adjusting the thresholds that are used to define whether or not results are considered statistically significant (6).

Trial Sequential Analysis is also able to assess when an intervention has an effect smaller than what would be considered clinically minimally important (6). Futility boundaries, originally developed for interim analysis in RCTs, can be estimated and used to provide a threshold below which an intervention would be considered to have no clinically important effect (6). Thus, performing further trials is considered futile as the intervention does not possess the postulated clinically minimally important effect (6).

In Trial Sequential Analysis, when neither the monitoring boundaries nor the futility boundaries are crossed, further information is usually required. Trial Sequential Analysis can also inform how much more information is required to get a conclusive answer regarding the effect of the intervention versus its comparator – this is called the distance between the accrued information and the required information.

Required information size:

For RCTs, an estimation of the required sample size is performed to ensure the number of participants included is enough to detect or reject a minimum clinically important effect size (17). For binary outcomes, such as death, the sample size estimation is based on the expected proportion of deaths in the control group, the expected relative risk reduction of the intervention, and the selected maximum risks of both type I and type II errors (18). Similarly, for meta-analyses to produce adequately powered findings regarding intervention efficacy, sufficient numbers of participants need to be included. This number is referred to as the 'required information size' (or 'optimal information size' or 'meta-analytic sample size') (22, 23). The meta-analytic required information size can be estimated using similar parameters as those used in sample size estimation for a single trial if one uses a fixed-effect model. If one intends to use a random-effects model, then one needs to consider adjusting for any between-study heterogeneity measured by inconsistency (I^2) or diversity (D^2) (18). Inconsistency is the test statistic for heterogeneity usually used in meta-analysis, and diversity characterises the proportion of between trial variation in any meta-analysis relative to the total model variance of the included trials (24). Diversity is equal to inconsistency or larger (24). Heterogeneity between studies is likely to be observed in meta-analyses due to the magnitude of the intervention effect varying when used in different study populations, in studies with different methodological characteristics, or due to variations in the intervention itself (13). Thus, sample size estimations need to be increased to allow for this between-trial heterogeneity (18).

In the Trial Sequential Analysis software, trials are chronologically ordered, and interim analyses are conducted as each trial is added using summary data from each trial. In a Trial sequential analysis where the 'required information size' has not been

reached, the threshold for statistical significance is inflated to account for sparse data and multiple testing of the interim analyses using monitoring boundaries; thus, the 95% confidence interval is not providing coverage of the real uncertainty and the cut-off for determining statistical significance is below the usual nominal figure of 0.05 (18). Furthermore, the Trial Sequential Analysis software provides adjusted confidence intervals if the 'required information size' has not been reached, which we refer to as Trial Sequential Analysis-adjusted confidence intervals (18). Technical details regarding how monitoring boundaries, information size, and Trial Sequential Analysis-adjusted confidence intervals are calculated can be found elsewhere (6, 18).

In the worked examples below, we show how the Trial Sequential Analysis software can be used to estimate the sample size required for one or more new trials to add further data to a meta-analysis to provide more firm evidence for an intervention either having or not having the postulated effect.

Results

In this section, we provide an example of how Trial Sequential Analysis successfully used data from feasibility and pilot RCTs that tested MiQuit, a text-message, self-help smoking cessation intervention for pregnant women, to justify research funds to undertake a third, more adequately powered RCT.

Previous MiQuit trials

Smoking during pregnancy increases the risk of miscarriage, stillbirth, low birth-weight, premature birth, perinatal morbidity and mortality, sudden infant death, as well as adverse infant behavioural outcomes (25, 26). Pregnancy is a life event which motivates cessation attempts amongst smokers and over 50% of pregnant women who smoke attempt to quit during this time (27), consequently pregnancy is an opportune moment to offer smoking cessation support. Text message, self-help support, smoking cessation programmes developed for non-pregnant smokers are effective, but such programmes are inappropriate for use during pregnancy (28-30). To address the lack of acceptable self-help, support cessation programmes for pregnant smokers in the UK, MiQuit was developed (7). MiQuit delivers individually-tailored text messages to pregnant smokers, with the aim of encouraging them to stop smoking (7). Further details on MiQuit can be found elsewhere (7).

A MiQuit feasibility RCT was conducted, including 207 women. Biochemically-validated, 7-day point prevalence cessation at 12 weeks post randomisation (~6 months gestation) was 12.5% in the experimental MiQuit group, compared with 7.8% in the control group (odds ratio (OR) 1.68, 95% confidence interval (CI) 0.66 to 4.31) (7). Although the trial was small, and the cessation period brief, the trial provided an estimate suggesting that MiQuit could have a positive impact in addition to routine care.

The feasibility RCT lead to minor changes to the intervention, before a pilot RCT was conducted to investigate the feasibility of undertaking a fully-powered multi-centre RCT in UK National Health Service (NHS) settings (8). The pilot MiQuit RCT recruited 407 pregnant women that smoke, which had largely similar baseline characteristics to those in the feasibility RCT. The self-reported abstinence from 4 weeks post-randomisation until late pregnancy follow-up (approximately 36 weeks gestation) biochemically validated at follow-up was 5.4% in the experimental MiQuit group versus 2.0% in the control group (OR 2.70, 95% CI 0.93 to 9.35) (8). This trial also suggested a beneficial effect of MiQuit.

As MiQuit is a cheap intervention and can be disseminated widely, it was anticipated that even a 1% to 2% absolute effect on smoking cessation in pregnancy could be clinically important and cost effective (8). The results from the feasibility and pilot trials suggested that an impact of this size was attainable; however, an adequately powered RCT would still be needed to determine whether MiQuit is effective and guide future routine clinical practise.

Conventional meta-analysis

The conventional way to determine if an intervention is effective or not is to use the naïve alpha of 5% and the naïve 95% confidence interval (10). Since both the feasibility and pilot trials used almost the same design as was planned to be used in the new RCT, they can be considered as pilots and it would be appropriate to meta-analyse these trials' findings together. Using a

random-effects model, a traditional meta-analysis of pilot and feasibility studies' data found, that women randomised to MiQuit were more than twice as likely to be abstinent in their pregnancy (pooled OR 2.26, 95% CI 1.04 to 4.93; I^2=0%, p=0.041). This result seems to be significant according to conventional assessment (p<0.05). However, this result should be interpreted with caution because, as described above, findings from meta-analyses based on only two small RCTs can produce spurious findings due to type I error (11, 12, 22) (please see below).

In the next sections, we use conventional sample size estimation methods to estimate the sample size for an RCT which, on its own would have enough power to show whether MiQuit might be effective, using a plausible treatment effect estimate derived from the conventional meta-analysis above. We also calculate a second sample size estimate for one or more further RCTs, which when pooled with data from feasibility and pilot trials using Trial Sequential Analysis methods, would be similarly decisive.

Conventional sample size estimation

As the pilot trial (8) was considered at lower risk of bias compared to the feasibility trial (7), a traditional sample size calculation using smoking cessation rate estimates derived from the pilot trial suggests a new trial would require a total sample size of 1292 participants. This estimate has 90% power (10% type II error) and 5% significance (2-sided test; type I error) to detect a 3.4% absolute difference in prolonged abstinence from smoking from 4 weeks after enrolment until 36 weeks gestation between the MiQuit and control groups (5.4% versus 2.0%) (8).

Trial Sequential Analysis

Figure 1.I illustrates a Trial Sequential Analysis incorporating findings from the MiQuit feasibility (A) (7) and pilot (B) (8) trials. In this Trial Sequential Analysis output, the x-axis represents the number of participants and marked on this are the numbers of participants recruited to each trial. The y-axis represents the z-score, where a positive z-score favours the MiQuit intervention and a negative z-score favours the control.

The z-score is the test that helps you decide whether to accept or reject the null hypothesis. Very high positive or very low negative z-scores are associated with very small p-values. The critical z-score values when using a 95% confidence level, which are known as the 'conventional test boundaries', are -1.96 and +1.96 and these relate to a two-sided p-value of 0.05. If the z-score is between -1.96 and +1.96, the p-value will be larger than 0.05, and the null hypothesis of no difference between intervention groups is accepted. The z-curve represents the cumulative z-score as each RCT is added to the analysis. In Figure 1.I, when trial B is added to the analysis, the z-curve crosses the conventional test boundary (p=0.05). This is consistent with the results from the conventional meta-analysis for MiQuit, where we found p=0.041.

The required information size is represented by the vertical red line in Figure 1. The required information size was estimated using the same variables as used for the conventional sample size estimation above (90% power, 5% significance, to detect a 3.4% absolute difference) (8); although this estimate could take into account observed heterogeneity, there was none in this meta-analysis (I^2 = 0% and D^2 = 0). Consequently, the estimated required information size of 1296 participants is only slightly different to that using conventional sample size estimation due to rounding errors. The estimate would be larger if heterogeneity were present.

As the cumulative z-curve does not cross the upper trial sequential monitoring boundary for benefit, this Trial Sequential Analysis shows that further information is required before any firm conclusion can be reached about MiQuit efficacy. Although the conventional meta-analysis suggested, with borderline significance, that pregnant women randomised to MiQuit were more than twice as likely to be abstinent from smoking in late pregnancy, the Trial Sequential Analysis software shows that this finding is not sufficiently robust. The Trial Sequential Analysis-adjusted confidence intervals for cessation using MiQuit (pooled OR 2.26, Trial Sequential Analysis-adjusted CI 0.66 to 7.70), are much wider than those of the conventional meta-analysis (pooled OR 2.26, 95% CI 1.04 to 4.93).

Without Trial Sequential Analysis having been undertaken, an interpretation of the conventional meta-analysis would have been that MiQuit is effective. However, Trial Sequential Analysis indicates that one cannot be secure in this interpretation and further trial data should be collected to eliminate the possibility that this is a false positive result, which can occur early in intervention evaluation when small trials are undertaken.

Calculating sample size for a third MiQuit RCT

Trial Sequential Analysis has demonstrated that further RCT data are required before a firm conclusion about MiQuit efficacy can be determined. As the initial two trials were sufficiently similar to be combined in Trial Sequential Analysis, we will now demonstrate how Trial Sequential Analysis can be used to estimate the sample size for (a) further trial(s) – data from which, when combined with the previous two trials in the Trial Sequential Analysis software, would be expected to provide a more decisive answer regarding MiQuit efficacy. We will also demonstrate how exemplar theoretical findings from future trials which are both in favour and against MiQuit having a positive effect would impact the Trial Sequential Analysis result.

Trial Sequential Analysis sample size estimation: Estimates derived from the Trial Sequential Analysis found the required information size as 1296 participants. From the feasibility and pilot studies, 605 women have already been recruited and randomised; therefore, the required sample size for further RCTs can be estimated as the difference between the required information size minus the number of women already recruited into the previous trials; thus a sample size of 691 women (346 per intervention group) would be needed, assuming a 1:1 ratio.

Figure 1.II shows the Trial Sequential Analysis output after adding a theoretical third trial (C) with a sample size of 630 women (315 per trial group), where an absolute difference of 3.17% was observed in favour of the MiQuit group versus the control group. The Trial Sequential Analysis clearly shows the cumulative z-curve line crossing the upper trial sequential monitoring boundary which indicates MiQuit being effective. As the trial sequential monitoring boundary has been crossed, the Trial Sequential Analysis z-curve does not need to reach the required information size of 1296. In the present scenario, we can firmly conclude that MiQuit is effective for smoking cessation compared with control (provided that all trials are valid and not influenced by systematic errors (bias) or other errors).

When a theoretical third trial (D) with a negative outcome is included in the Trial Sequential Analysis (figure 1.III), we observe a different output. Here, the third trial of sample size 630 was intentionally given a negative outcome (absolute difference of -0.63% in favour of control). Here we observe the z-curve drop below the conventional test boundary, and in a meta-analysis we would have concluded that MiQuit was not effective. However, in the Trial Sequential Analysis, the futility boundary is not crossed, so we are unable to decisively say that MiQuit is not as effective as control for smoking cessation. Due to the diversity, the required information size has increased to 1941, meaning future trials will need a further 706 participants.

A conservative approach to sample size estimation using Trial Sequential Analysis: In the above example, the required information size was derived using the smoking cessation effect from the pilot trial (8). Therefore, it can be contested whether data from the pilot trial should be included in subsequent Trial Sequential Analysis. Consequently, one could exclude the data from the pilot trial from the Trial Sequential Analysis and re-estimate the total number required (figure 2.I). Using this approach, to provide a conclusive result, either a single trial of 1098 participants (549 per intervention group, assuming a 1:1 ratio) or multiple trials cumulating to a total of 1098 participants, would be needed. This figure, although conservative, is still less than the estimate from the conventional sample size calculation.

Figures 2.II and 2.III also show the Trial Sequential Analysis outputs if theoretical trials C and D were included in the analyses. In both situations further information is needed, despite the z-curve coming close to the upper trial sequential monitoring boundary in figure 2.II and the futility boundary in figure 2.III.

Sensitivity analysis

The modelled scenario, in which there is no heterogeneity between trials in a meta-analysis is rare; in most situations where the described approach is used, some heterogeneity between studies is to be expected. Trial Sequential Analysis provides 95% confidence intervals for heterogeneity (D^2) within meta-analyses. One way to fully allow for heterogeneity is to perform a sensitivity analysis using the upper 95% confidence interval for the between-trial heterogeneity variance estimate. This would increase the required information size. In our example, the program could not calculate the 95% confidence interval surrounding the D^2 of 0% as there were less than three included studies. In this case it is possible to input an estimate for heterogeneity into the Trial Sequential Analysis software.

Discussion

The above example demonstrates how Trial Sequential Analysis can be used to determine the required sample size for one or more additional RCTs to make a meta-analysis more conclusive. This sample size would be considered underpowered in comparison to a traditional RCT sample size calculation. By using Trial Sequential Analysis in such a way, future trials could be planned using significantly fewer resources and with less cost than trials planned using traditional sample size calculations.

In the worked example, data from the pilot trial were used in the Trial Sequential Analysis to estimate the required information size. Ignoring that the same data is being used twice (for the estimation and for the meta-analysis) could mean that the estimate generated is not sufficiently conservative. Thus, we present a modification which attempts to overcome this issue. This approach increases the difference between required information size minus the accrued information by the sample size of the trial used in the estimation.

It is important to note that in the example, the meta-analysis of the existing two MiQuit trials quantified heterogeneity as 0%, indicating no heterogeneity. However, it is unlikely that this will be the case for meta-analyses of other interventions aimed at changing addictive behaviours (31, 32); therefore, trial sequential analysis methods have been developed to account for this (22). In Trial Sequential Analysis, estimated information size and monitoring boundaries, vary with the level of heterogeneity in the meta-analysis, the greater the level of heterogeneity, the larger the sample size and the wider the monitoring boundaries needed to reach firm conclusions about the effectiveness of the intervention. This is because the required information size is calculated relative to the measure of heterogeneity, the fraction of the accrued information size and the point estimate (18).

In the examples presented, odds ratios were also used instead of relative risk, as the feasibility study was powered using an odds ratio from a meta-analysis investigating mobile phone interventions for smoking cessation in the general population (7). Moreover, the quit rates are relatively low, so there is very little difference between the odds ratio and relative risk. In other trial sequential analyses, it may be advisable to use relative risks instead of odds ratios, to avoid overestimates. Additionally, it may be inappropriate to use the odds ratio used to power the feasibility trial to estimate sample sizes for future MiQuit trials since data now exists from the feasibility and pilot trials. In our example, the stipulated intervention effect was derived from the pilot trial ('internal data'), and it may be argued that such adaptive data should not be used in meta-analysis (33).

Kulinskaya and Wood argued that in an underpowered meta-analysis, not only is it necessary to assess the gap from the accrued information size to the required information size (i.e. the number of additional participants you need to randomise), but also the number of trials that should be conducted to randomise this number of participants (34). Using multiple trials to reach the required information size may be beneficial in meta-analyses where heterogeneity occurs (34). Smaller trials have more imprecise estimates of intervention effects; hence heterogeneity is reduced in the meta-analysis of such trials. However, setting up more than one trial can be more expensive and may not be realistic in practice.

Recently, the Cochrane Collaboration evaluated and updated their guidance on using sequential approaches in meta-analysis in their reviews (5, 10, 35). The Cochrane Handbook authors concluded that sequential methods should not be used in primary analyses or to draw conclusions, but could be used as secondary analyses in reviews if they are prospectively planned and the assumptions underlying the design are clearly justified (5, 10). In their guidance, the evidence synthesis group state that authors' interpretations of evidence should be based on estimated magnitude of effect of an intervention and its uncertainty rather than drawing binary conclusions, and decisions should not be influenced by plans for future updates of meta-analyses (10). These

criticisms of sequential approaches in meta-analyses apply to the traditional use of Trial Sequential Analysis, whereas our paper demonstrates an alternative use of the method.

Another reason given by The Cochrane Handbook authors against using sequential methods as a primary analysis in reviews, is the argument that a meta-analyst does not have any control over designing trials that are eligible for meta-analysis (10). It would therefore be impossible to construct a set of stopping rules (10). In our example, the opposite is the case. Both the feasibility and pilot trials were conducted by the same group of investigators, and any future trials would have a consideration for the desired properties of a stopping rule.

Finally, The Cochrane Handbook authors also highlight that there are methodological limitations to sequential methods when heterogeneity is present (10). In the example described in this paper, heterogeneity was not detected, possibly due to the lack of sufficient power to detect a moderate level. However, we do discuss how the presence of heterogeneity can be overcome in Trial Sequential Analysis by performing a sensitivity analysis.

Conclusions

In conclusion, Trial Sequential Analysis is a freely available software that can utilise data from feasibility and pilot trials as well as other trials, in order to estimate a sample size for one or more future RCTs, to provide an adequately powered conclusion regarding an intervention's benefits and harms. This simple use of expensively collected trial data could be usefully exploited by researchers evaluating other interventions.

Abbreviations

CI – Confidence interval

NHS – National Health Service

OR - Odds ratio

RCT – Randomised clinical trial

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Trial Sequential Analysis software, user manual and further information regarding the mathematics behind the method are available at http://www.ctu.dk/tsa/ for free.

All data generated or analysed during this study are included in the following published articles:

Naughton F, Prevost AT, Gilbert H, Sutton S. Randomized controlled trial evaluation of a tailored leaflet and SMS text message self-help intervention for pregnant smokers (MiQuit). Nicotine & Tobacco Research. 2012;14(5):569-77.

Naughton F, Cooper S, Foster K, Emery J, Leonardi-Bee J, Sutton S, et al. Large multi-centre pilot randomized controlled trial testing a low-cost, tailored, self-help smoking cessation text message intervention for pregnant smokers (MiQuit). Addiction. 2017;112(7):1238-49.

Competing interests

RC, CG, IB and TC declare that they have no competing interests.

JLB reports fees from undertaking independent statistical review for Danone Nutricia Research, and in relation to providing statistical expertise to the Food Standards Agency, both outside the subject of the submitted work.

Funding

This study is funded by the National Institute for Health Research (NIHR) Applied Research Collaboration East Midlands (ARC EM). Professor Coleman is a NIHR Senior Investigator. The views expressed are those of the author(s) and not necessarily those of the NIHR, the Department of Health and Social Care, or Rigshospitalet.

Authors' contributions

RC, JLB, IB and TC conceived the idea for this manuscript. RC input all data into the software and produced the results. RC, JLB and CG all contributed to the interpretation of the data. RC produced an initial draft of the manuscript, and all authors made substantial revisions to the work. All authors commented on the final draft of the manuscript and RC finalised the text. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Arain M, Campbell MJ, Cooper CL, Lancaster GA. What is a pilot or feasibility study? A review of current practice and editorial policy. BMC Medical Research Methodology. 2010;10(1):67.
2. Wittes J, Brittain E. The role of internal pilot studies in increasing the efficiency of clinical trials. Statistics in Medicine. 1990;9(1-2):65-72.
3. Thabane L, Ma J, Chu R, Cheng J, Ismaila A, Rios LP, et al. A tutorial on pilot studies: the what, why and how. BMC Medical Research Methodology. 2010;10(1):1.
4. Brok J, Thorlund K, Gluud C, Wetterslev J. Trial sequential analysis reveals insufficient information size and potentially false positive results in many meta-analyses. J Clin Epidemiol. 2008;61(8):763-9.
5. Thomas J, Askie L, Berlin J, Elliott J, Ghersi D, Simmonds M, et al. Chapter 22: Prospective approaches to accumulating evidence: Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019); 2019 Available from: www.training.cochrane.org/handbook.
6. Thorlund K, Engstrøm J, Wetterslev J, Brok J, Imberger G, Gluud C. User manual for trial sequential analysis (TSA). Copenhagen Trial Unit, Centre for Clinical Intervention Research, Copenhagen, Denmark. 2011;1:1-115.
7. Naughton F, Prevost AT, Gilbert H, Sutton S. Randomized controlled trial evaluation of a tailored leaflet and SMS text message self-help intervention for pregnant smokers (MiQuit). Nicotine & Tobacco Research. 2012;14(5):569-77.
8. Naughton F, Cooper S, Foster K, Emery J, Leonardi-Bee J, Sutton S, et al. Large multi-centre pilot randomized controlled trial testing a low-cost, tailored, self-help smoking cessation text message intervention for pregnant smokers (MiQuit). Addiction. 2017;112(7):1238-49.

9. Garattini S, Jakobsen JC, Wetterslev J, Bertelé V, Banzi R, Rath A, et al. Evidence-based clinical practice: Overview of threats to the validity of evidence and how to minimise them. European Journal of Internal Medicine. 2016;32:13-21.
10. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. *Cochrane Handbook for Systematic Reviews of Interventions*: Cochrane; 2019. Available from: www.training.cochrane.org/handbook.
11. Imberger G, Thorlund K, Gluud C, Wetterslev J. False-positive findings in Cochrane meta-analyses with and without application of trial sequential analysis: an empirical review. BMJ Open. 2016;6(8):e011890.
12. Thorlund K, Imberger G, Walsh M, Chu R, Gluud C, Wetterslev J, et al. The number of patients and events required to limit the risk of overestimation of intervention effects in meta-analysis—a simulation study. PLoS ONE. 2011;6(10):e25491.
13. Imberger G, Gluud C, Boylan J, Wetterslev J. Systematic reviews of anesthesiologic interventions reported as statistically significant: problems with power, precision, and type 1 error protection. Anesthesia & Analgesia. 2015;121(6):1611-22.
14. Harrison W, Angoulvant F, House S, Gajdos V, Ralston SL. Hypertonic Saline in Bronchiolitis and Type I Error: A Trial Sequential Analysis. Pediatrics. 2018;142(3):e20181144.
15. Simmonds M, Salanti G, McKenzie J, Elliott J, Agoritsas T, Hilton J, et al. Living systematic reviews: 3. Statistical methods for updating meta-analyses. J Clin Epidemiol. 2017;91:38-46.
16. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. PLOS Med. 2007;4(3):e78.
17. Jackson D, Turner R. Power analysis for random-effects meta-analysis. Research Synthesis Methods. 2017;8(3):290-302.
18. Wetterslev J, Jakobsen JC, Gluud C. Trial Sequential Analysis in systematic reviews with meta-analysis. BMC Medical Research Methodology. 2017;17(1):39.
19. Brok J, Thorlund K, Wetterslev J, Gluud C. Apparently conclusive meta-analyses may be inconclusive—trial sequential analysis adjustment of random error risk due to repetitive testing of accumulating data in apparently conclusive neonatal meta-analyses. International Journal of Epidemiology. 2009;38(1):287-98.
20. Nguyen T-L, Collins GS, Lamy A, Devereaux PJ, Daurès J-P, Landais P, et al. Simple randomization did not protect against bias in smaller trials. J Clin Epidemiol. 2017;84:105-13.
21. Borm GF, Donders ART. Updating meta-analyses leads to larger type I errors than publication bias. J Clin Epidemiol. 2009;62(8):825-30.
22. Wetterslev J, Thorlund K, Brok J, Gluud C. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. J Clin Epidemiol. 2008;61(1):64-75.
23. Pogue JM, Yusuf S. Cumulating evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis. Controlled clinical trials. 1997;18(6):580-93.
24. Wetterslev J, Thorlund K, Brok J, Gluud C. Estimating required information size by quantifying diversity in random-effects model meta-analyses. BMC Medical Research Methodology. 2009;9(1):86.
25. Batstra L, Hadders-Algra M, Neeleman J. Effect of antenatal exposure to maternal smoking on behavioural problems and academic achievement in childhood: Prospective evidence from a Dutch birth cohort. Early Hum Dev. 2003;75(1-2):21-33.
26. Turner-Warwick M. Smoking and the Young: A report of a working party of the Royal College of Physicians. Tobacco Control. 1992;1(3):231-5.
27. McAndrew F, Thompson J, Fellows L, Large A, Speed M, Renfrew MJ. Infant Feeding Survey 2010: Health and Social Care Information Centre: Health and Social Care Information Centre; 2012 [Available from: http://www.hscic.gov.uk/catalogue/PUB08694/Infant-Feeding-Survey-2010-Consolidated-Report.pdf http://digital.nhs.uk/catalogue/PUB08694.]
28. Abroms LC, Ahuja M, Kodl Y, Thaweethai L, Sims J, Winickoff JP, et al. Text2Quit: results from a pilot test of a personalized, interactive mobile health smoking cessation program. Journal of Health Communication. 2012;17(sup1):44-53.
29. Abroms LC, Boal AL, Simmens SJ, Mendel JA, Windsor RA. A randomized trial of Text2Quit: a text messaging program for smoking cessation. American Journal of Preventive Medicine. 2014;47(3):242-50.
30. Free C, Whittaker R, Knight R, Abramsky T, Rodgers A, Roberts IG. Txt2stop: a pilot randomised controlled trial of mobile phone-based smoking cessation support. Tobacco Control. 2009;18(2):88-91.

31. Higgins JPT. Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. International Journal of Epidemiology. 2008;37(5):1158-60.
32. Thorlund K, Imberger G, Johnston BC, Walsh M, Awad T, Thabane L, et al. Evolution of heterogeneity (I2) estimates and their 95% confidence intervals in large meta-analyses. PLOS ONE. 2012;7(7):e39471.
33. Bauer P, Bretz F, Dragalin V, König F, Wassmer G. Twenty-five years of confirmatory adaptive designs: opportunities and pitfalls. Statistics in Medicine. 2016;35(3):325-47.
34. Kulinskaya E, Wood J. Trial sequential methods for meta-analysis. Research Synthesis Methods. 2014;5(3):212-20.
35. Schmid C, Senn S, Sterne J, Kulinskaya E, Posch M, Roes K, et al. Should Cochrane apply error-adjustment methods when conducting repeated meta-analyses? : Cochrane Scientific Committee; 2018 [Available from: https://methods.cochrane.org/sites/default/files/public/uploads/tsa_expert_panel_guidance_and_recommendation_final.pdf.

Appendix C

Professional Development

Masters modules

Advanced Statistical Methods (9 training units)

Research Methods in Epidemiology with Basic Statistics (19 training units)

Tobacco control interventions (9 training units)

Graduate school short courses

Microsoft Word: Creating and Managing Long Documents (2 training units)

Thinking Ahead - Career options and planning for PhDs (1 training unit)

Preparing for your confirmation review (0 training units)

Problems with academic writing (0.5 training units)

Structuring Your Thesis (1 training unit)

Faculty postgraduate Research Forum (Medicine and Health Sciences Faculty) (4 training units)

Advanced presentation skills for researchers (moderated online learning course) (2 training units)

Structuring Your Thesis (1 training unit)

Drafting a Chapter of your Thesis (1 training unit)

Applying for academic jobs - PhD students (1 training unit)

Applying for jobs outside academia - PhD students (1 training unit)

Editing Academic Writing (0.5 training units)

Preparing for the viva (1 training unit)

Creating a strong argument for your thesis (0.5 training units)

External courses

Joanna Briggs Institute (JBI) Comprehensive Systematic Review Training Program (10 training units)

NIHR CLAHRC EM skills session: Preparing your data for analysis with STATA

NIHR CLAHRC EM skills session: Writing for publication

Conferences and seminars

University of Nottingham Medicine & Health Sciences Faculty Postgraduate Research Forum 2017 – '2 minutes of impact' pitch and printed poster: Nicotine replacement therapy for smoking cessation during pregnancy. (4 training units)

Society for Research on Nicotine and Tobacco (SRNT); Florence, Italy 2017 – Poster presentation: Nicotine Replacement Therapy for Smoking Cessation during Pregnancy: A Trial Sequential Analysis.

SRNT-Europe; Munich, Germany 2018 – Poster presentation: Saliva cotinine concentrations in pregnant women who smoke whilst using nicotine replacement therapy

11th Congrès national de la Société Francophone de Tabacologie (CSFT; National meeting of the Francophone Society of Tobacco); Paris, France 2017 – Attendance only.

12th CSFT; Montpellier, France 2018 – Oral presentation: Mobile phone interventions for smoking cessation in pregnant women that smoke. This presentation incorporated some of the findings in **Chapter 4**.

NIHR ARC EM 3-Minute Thesis Presentation Day; Nottingham, UK 2018 – Oral presentation: NRT for smoking cessation during pregnancy.

University of Nottingham Sue Watson Oral Presentation Event; Derby, UK 2019 – Oral presentation: Saliva cotinine levels in pregnant women who smoke and use nicotine patches. (2 training units)

I have performed oral presentations describing TSA at research seminars for the Division of Primary Care, and Epidemiology and Public Health at the University of Nottingham; as well as at research seminars for biostatistics and pharmacology at the Pitié-Salpêtrière hospital in Paris, France.

Prizes and awards

Awarded £470.17 from the School of Medicine Doctoral Programmes Committee Support Fund – used towards travel for SRNT 2017.

Appendix D

Trial Sequential Analysis software

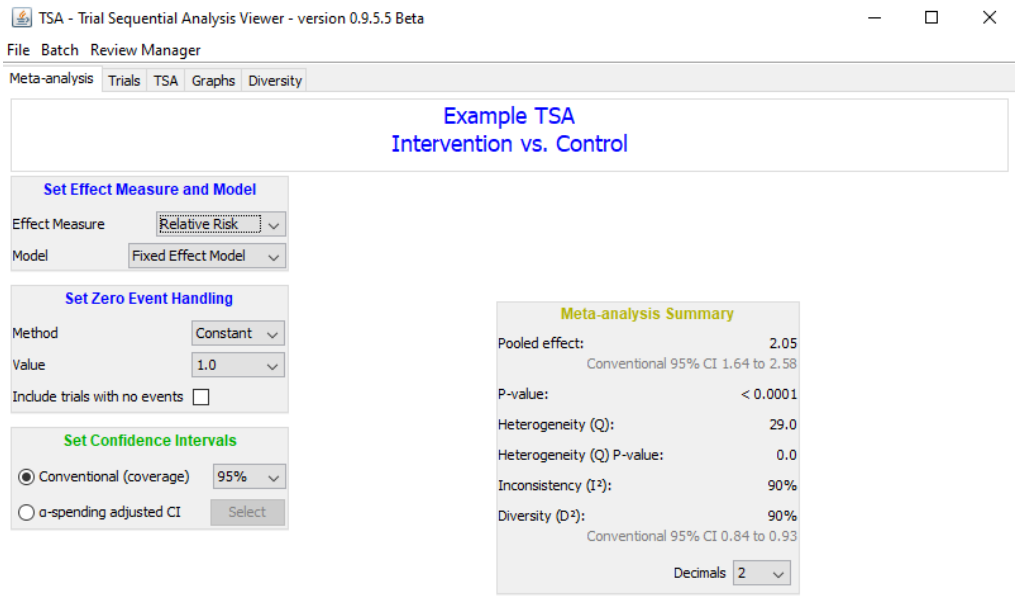

Figure S7 Meta-analysis summary page. Here you can change the effect measure (relative risk or odds ratio), and the model (random effects or fixed effects model).

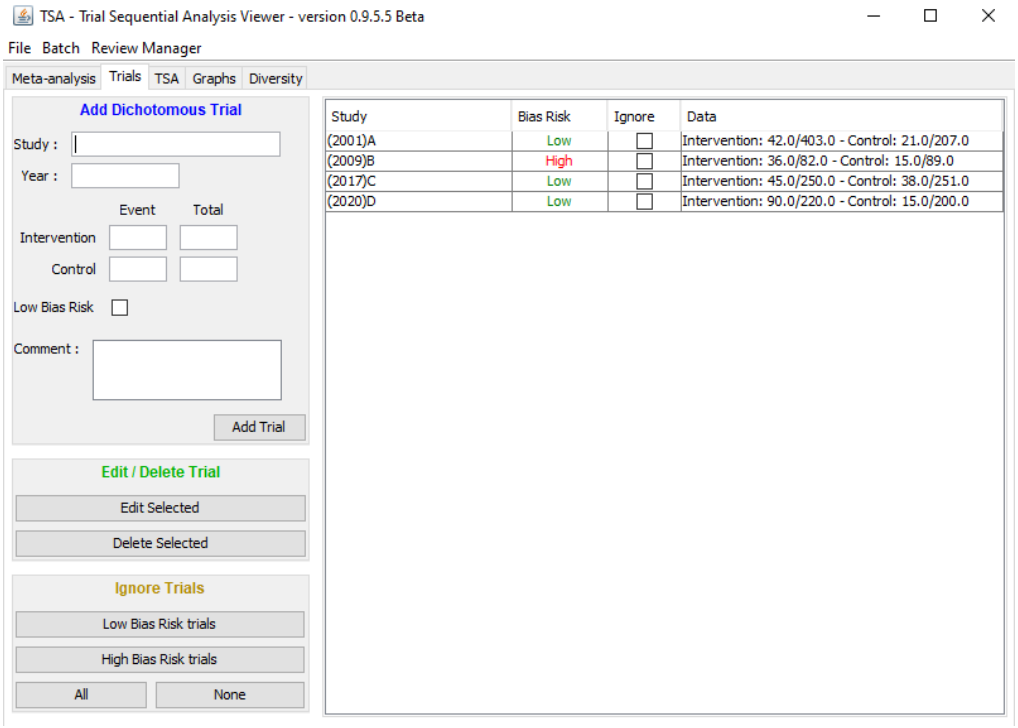

Figure S8 Included trials. Here you include all trials to be included in the TSA, and their results.

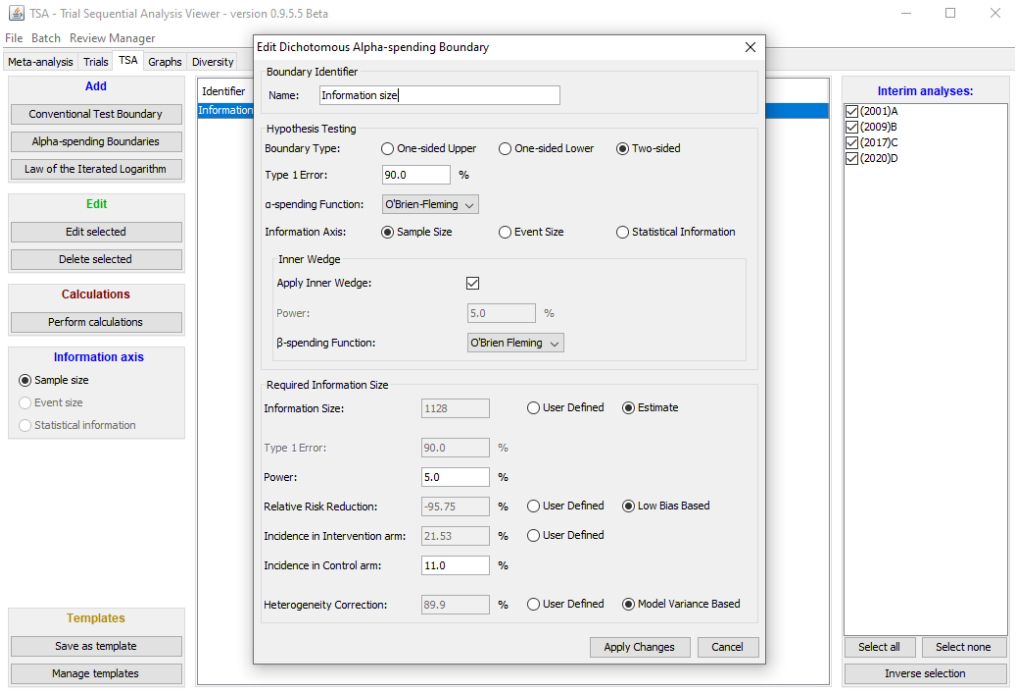

Figure S9 Setting the parameters to calculate the information size and monitoring boundaries. Here you input *a priori* assumptions for the TSA. Once all relevant information is input the TSA will calculate the information size and monitoring boundaries, and create the TSA graph.